

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:46:01 ; Search time 28.95 Seconds
(without alignments)
285,176 Million cell updates/sec

Title: US-09-980-881-3
Perfect score: 1798
Sequence: 1 FOSGQVLAALPRISROYQVLT.....IKYSFTSNPVEKILPLSLK 338

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 segs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1699.5	94.5	423	1	US-07-649-591B-3
2	1699.5	94.5	423	1	US-08-277-540-3
3	1699.5	94.5	423	1	US-08-430-787A-3
4	1699.5	94.5	423	2	US-08-869-057-2
5	646.5	36.0	404	1	US-08-696-139-2
6	637.5	35.5	415	2	US-08-860-882A-57
7	632	35.2	417	1	US-07-649-591B-7
8	632	35.2	417	1	US-08-277-540-7
9	632	35.2	417	1	US-08-430-787A-7
10	622	34.6	417	1	US-07-649-591B-6
11	622	34.6	417	1	US-08-277-540-6
12	622	34.6	417	1	US-08-430-787A-6
13	619.5	34.5	716	4	US-09-171-945-125
14	610	33.9	307	2	US-08-782-760-6
15	610	33.9	307	5	PCT-US96-00995-6
16	609.5	33.9	396	1	US-07-649-591B-4
17	609.5	33.9	396	1	US-08-277-540-4
18	609.5	33.9	396	1	US-08-430-787A-4
19	582	32.4	306	1	US-08-696-139-4
20	549.5	30.6	417	1	US-07-649-591B-8
21	549.5	30.6	417	1	US-08-277-540-8
22	549.5	30.6	417	1	US-08-430-787A-8
23	549	30.5	613	4	US-09-171-945-113
24	532.5	29.6	419	1	US-07-649-591B-5
25	532.5	29.6	419	1	US-08-277-540-5
26	532.5	29.6	419	1	US-08-430-787A-5
27	516.5	28.7	417	4	US-08-640-906-4

28	516.5	28.7	417	4	US-09-395-936-4	Sequence 4, Appl
29	515.5	28.7	417	4	US-08-640-906-18	Sequence 18, Appl
30	515.5	28.7	417	4	US-09-395-936-18	Sequence 18, Appl
31	491.5	27.3	419	4	US-08-640-906-2	Sequence 2, Appl
32	491.5	27.3	419	4	US-09-395-936-2	Sequence 2, Appl
33	486.5	27.1	419	4	US-08-640-906-17	Sequence 17, Appl
34	486.5	27.1	419	4	US-09-395-936-17	Sequence 17, Appl
35	167	9.3	37	1	US-07-649-591B-1	Sequence 1, Appl
36	167	9.3	37	1	US-08-277-540-1	Sequence 1, Appl
37	167	9.3	37	1	US-08-430-787A-1	Sequence 1, Appl
38	116.5	6.5	706	1	US-08-484-105-16	Sequence 16, Appl
39	116.5	6.5	706	1	US-08-484-106-16	Sequence 16, Appl
40	102	5.7	734	4	US-08-706-216-2	Sequence 2, Appl
41	97	5.4	95	2	US-08-782-760-3	Sequence 3, Appl
42	97	5.4	95	5	PCT-US96-00995-3	Sequence 3, Appl
43	94	5.2	1149	1	US-07-915-203-2	Sequence 2, Appl
44	94	5.2	1149	1	US-08-272-887-2	Sequence 2, Appl
45	94	5.2	1149	2	US-08-789-449-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 5.5e-175;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
DB 23 FOSGQVLAALPRISROYQVLTFTTEYEIVLMQVPTADLVKKQVHFFVNASQDVNKA 82

QY 61 HLNVSGLPCSVLLADVEDLIIQOQISNDTVSPRASASYEQYHSLNEITYSWIEFTERRHPD 120
DB 83 HLNVSGLPCSVLLADVEDLIIQOQISNDTVSPRASASYEQYHSLNEITYSWIEFTERRHPD 142
QY 121 MLTKIHGSSFEKPYLYLVAKVSGKEQTAKNAIWDGCIHAREWISPAFCWFIH----- 175
DB 143 MLTKIHGSSFEKPYLYLVAKVSGKEQTAKNAIWDGCIHAREWISPAFCWFIHITQFY 202
QY 176 -----NRMWRKRSFYANNHCIGTDLNRNPAK 203
DB 203 GIIQYTNLRLVDYVMPVYVNDGYDYSWKNNRKRNSFYANNHCIGTDLNRNPAK 262
QY 204 HMCBEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIPEPSY 263
DB 263 HMCBEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIPEPSY 322
QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRTYTHGSETLYLAPGGDDWYIDLGIRYSF 323
DB 323 TRSKSDHEELSLVASEAVRAIEKTSKNTRTYTHGSETLYLAPGGDDWYIDLGIRYSF 382
QY 324 T 324
DB 383 T 383

RESULT 2
US-08-277-540-3
Sequence 3, Application US/08277540
Patent No. 5474901
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 5,5e-175;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 FOSGOYLAALPRTSRVOYVLONTTTEYELVMOPTADLIYKKQVHFFVNASDVDNKA 60
DB 23 FOSGOYLAALPRTSRVOYVLONTTTEYELVMOPTADLIYKKQVHFFVNASDVDNKA 82
QY 61 HLNVSGLPCSVLLADVEDLIIQOQISNDTVSPRASASYEQYHSLNEITYSWIEFTERRHPD 120
DB 83 HLNVSGLPCSVLLADVEDLIIQOQISNDTVSPRASASYEQYHSLNEITYSWIEFTERRHPD 142
QY 121 MLTKIHGSSFEKPYLYLVAKVSGKEQTAKNAIWDGCIHAREWISPAFCWFIH----- 175
DB 143 MLTKIHGSSFEKPYLYLVAKVSGKEQTAKNAIWDGCIHAREWISPAFCWFIHITQFY 202
QY 176 -----NRMWRKRSFYANNHCIGTDLNRNPAK 203
DB 203 GIIQYTNLRLVDYVMPVYVNDGYDYSWKNNRKRNSFYANNHCIGTDLNRNPAK 262
QY 204 HMCBEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIPEPSY 263
DB 263 HMCBEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIPEPSY 322
QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRTYTHGSETLYLAPGGDDWYIDLGIRYSF 323
DB 323 TRSKSDHEELSLVASEAVRAIEKTSKNTRTYTHGSETLYLAPGGDDWYIDLGIRYSF 382
QY 324 T 324
DB 383 T 383

RESULT 3
US-08-430-787A-3
Sequence 3, Application US/08430787A
Patent No. 5593674
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616

REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 5.5e-175;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

OY 1 FOSGOVLALPRTSROVOYLQNLTTTTEIYVLMOPVTADLIYKKQVHFVNASDVNVKA 60
DB 23 FOSGOVLALPRTSROVOYLQNLTTTTEIYVLMOPVTADLIYKKQVHFVNASDVNVKA 82

OY 61 HLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNIEYSWIEFTTERHPD 120
DB 83 HLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNIEYSWIEFTTERHPD 142

OY 121 MLTKIHGSSFEKPYLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCIMFIGH----- 175
DB 143 MLTKIHGSSFEKPYLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCIMFIGHITQFY 202

OY 176 -----NRMWRKRSFYANNHCIGTDINRNFASK 203
DB 203 GIIGQYTNLRLVDFYVMPVNVGDYDYSWKKNRMWRKRSFYANNHCIGTDINRNFASK 262

OY 204 HMCBEGASSSCSEYTCGLPESPEVKAVASFLRRNINQIKAYISMHSYQHIIVPEYSY 263
DB 263 HMCBEGASSSCSEYTCGLPESPEVKAVASFLRRNINQIKAYISMHSYQHIIVPEYSY 322

OY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGIKYSF 323
DB 323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGIKYSF 382

OY 324 T 324
DB 383 T 383

RESULT 4
US-08-869-057-2
Sequence 2, Application US/08869057
Patent No. 5983562
GENERAL INFORMATION:
APPLICANT: Morser, Michael J
APPLICANT: Nagashima, Mariko
TITLE OF INVENTION: Method of Detecting Thrombotic Disease
TITLE OF INVENTION: Risk
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Berlex Biosciences Legal Department
STREET: 15049 San Pablo Avenue
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94804-0099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,057
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Washtien, Wendy L
REGISTRATION NUMBER: 36,301
REFERENCE/DOCKET NUMBER: 51509AUSM1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-262-5411
TELEFAX: 510-262-7095
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE: Plasma
TISSUE TYPE:
FEATURE:
NAME/KEY: Peptide
LOCATION: 23.401
US-08-869-057-2

Query Match 94.5%; Score 1699.5; DB 2; Length 423;
Best Local Similarity 89.8%; Pred. No. 5.5e-175;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

OY 1 FOSGOVLALPRTSROVOYLQNLTTTTEIYVLMOPVTADLIYKKQVHFVNASDVNVKA 60
DB 23 FOSGOVLALPRTSROVOYLQNLTTTTEIYVLMOPVTADLIYKKQVHFVNASDVNVKA 82

OY 61 HLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNIEYSWIEFTTERHPD 120
DB 83 HLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSLNIEYSWIEFTTERHPD 142

OY 121 MLTKIHGSSFEKPYLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCIMFIGH----- 175
DB 143 MLTKIHGSSFEKPYLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCIMFIGHITQFY 202

OY 176 -----NRMWRKRSFYANNHCIGTDINRNFASK 203
DB 203 GIIGQYTNLRLVDFYVMPVNVGDYDYSWKKNRMWRKRSFYANNHCIGTDINRNFASK 262

OY 204 HMCBEGASSSCSEYTCGLPESPEVKAVASFLRRNINQIKAYISMHSYQHIIVPEYSY 263
DB 263 HMCBEGASSSCSEYTCGLPESPEVKAVASFLRRNINQIKAYISMHSYQHIIVPEYSY 322

OY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGIKYSF 323
DB 323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDMIYDLGIKYSF 382

OY 324 T 324
DB 383 T 383

RESULT 5
US-08-696-139-2
Sequence 2, Application US/08696139
Patent No. 5672061
GENERAL INFORMATION:
APPLICANT: Fayerman, Jeffrey T.
APPLICANT: Greenen, David P.
APPLICANT: Hershteger, Charles L.
APPLICANT: Larson, Jeffrey L.
APPLICANT: Steiner, Jane L.
APPLICANT: Zhang, Haichao
TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana

APPLICANT: HENNAM, JOHN FREDERICK
 APPLICANT: BLAKELY, DAVID CHARLES
 APPLICANT: MARSHAM, PETER ROBERT
 APPLICANT: HEATON, DAVID WILLIAM
 APPLICANT: DAVIES, DAVID HUW
 TITLE OF INVENTION: CHEMICAL COMPOUNDS
 NUMBER OF SEQUENCES: 77
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PILLSBURY, MADISON & SUTRO
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy Disk
 COMPUTER: IBM compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/860,882A
 FILING DATE: JUNE 23, 1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: DONALD J. BIRD
 REGISTRATION NUMBER: 25,323
 REFERENCE/DOCKET NUMBER: 9901/238653

	Query Match	Similarity	Score	DB 2:	Length
	Best Local	36.0%	Pred. No.	2.6e-60;	
	Matches	130;	Conservative	68;	Mismatches 120; Indels 43; Gaps
Qy	1	FQSGGVLAALPRTSRQVQVLQNIITTYTEIVLQOPVYADLIYKKQVHFVYNASDVQNVKA	60		
	1				
Db	20	FEQGEKFRVNEDEENHINIIIRELASTQIDPFMKPSVQIQKHSHYDFEYKAEEDITYAZN	79		
Qy	61	HLNVSIPCSVLLADVEDLIDQOISNDIYSPASASIVYQVYSLSEIYSWIEFTTERHPD	120		
	80	VLKQNELQKTVLISNLRNVEAOFDSRV---RATGSHSEKYKKWETIEAWTQVATEMNA	136		
Db	121	MLTKIHISGFPEKYPLVYLKVGSKDQTKAIAWIDQIHARPMISPARCLMTI	173		
	137	LISRSVIGITTFEGRAIYLLK-GKAGQNKPALEFMCQGHARHWISPAFCQWVRAVATRY	195		
Qy	174	-----GH-----NRMRKRSFEYVANNHCIGTFLNENFSK	203		
	196	GREIQTVELDKLDEYVLPVLINDGYITWTWSRMRWTRSHIHGSSSICIGDPNNNF-DA	254		
Qy	204	HMCEGASSSSCSEFYCCLYPESFEYVAVASFLRLRNIOIKAYTSMHSYQOHIFPYSY	263		
	255	GWCEIGASRNPCDEYTCGPAASESEKETALADFINKLSIKAYLLTISYSQMMIYPSY	314		
Qy	264	TRSKSDHEELSLVASEVRAIEKTSKWTRTYHGHGSETLIAPCGGDWYIIDGIRKSF	323		
	315	AYKIGENNAELNALAKAIYKEL-ASLHGTKITTYGCAITTYIPAAGGSDDWAVDQIRYSF	373		
Qy	324	T 324			
	374	T 374			
Db					


```

QY 61 HUNVSGIPCSVLADVEDLIQOQIS-NDTVSPASASYEQYHSLNEIYSWIEFITERHP 119
DB 81 ALDQNMHEIILHDLQEEIEKQFVKEDIPGRHS--YAKNNMEKIYAMTEKMDKYP 137
QY 120 DMLTJHIGSSPEKPYLYLVKVGSKQETAKNAIMIDCGIHAREMISPACLMFI----- 173
DB 138 EWSRKIGISTVEDNPLVYKI-GEKNERKALFMDCGIHAREMWSPAFCQWVYQATKT 196
QY 174 -----GH-----NRMRKRNSFYANNHCIGTDLNNRFAS 202
DB 197 YGRNKIMTKLLDMNFIYILFVFNVDGYIMSWTRNMRKRNSKONSKCIGTDLNNRF-N 255
QY 203 KHWCEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIYPPYS 262
DB 256 ASWNSIPNTNDPCADNIRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYQMLFPPYG 315
QY 263 YTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHSFETLYLAPGGDDMIYDLGIKYS 322
DB 316 YTSKLPNNHEDLAKVAKIGTDVL-STREYETRYIYGPIESTIYISGSSLDMAWDLGIKHT 374
QY 323 F 323
DB 375 F 375

```

```

RESULT 11
US-08-277-540-6
: Sequence 6, Application US/08277540
: Patent No. 5474901
:
: GENERAL INFORMATION:
: APPLICANT: Drayna, Dennis T., Eaton, Dan L.
: TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patln (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/277,540
: FILING DATE: 19-JUL-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/167727
: FILING DATE: 15-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/959944
: FILING DATE: 14-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/649591
: FILING DATE: 01-FEB-91
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 688D1C1D1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 417 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
: US-08-277-540-6

```

```

Query Match 34.6%; Score 622; DB 1; Length 417;
Best Local Similarity 35.5%; Pred. No. 1,2e-58;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;

QY 1 FQSGVLAALPRTSROYVQIONTTTEIYELVQPVADLIYKKQVHFVNASDVNKA 60
DB 21 FDEKEFERYVPODEKQADILIKDLAKTNEIDFWYPGATHHVAAMMDVFRVSEKESQAIQS 80
QY 61 HUNVSGIPCSVLADVEDLIQOQIS-NDTVSPASASYEQYHSLNEIYSWIEFITERHP 119
DB 81 ALDQNMHEIILHDLQEEIEKQFVKEDIPGRHS--YAKNNMEKIYAMTEKMDKYP 137
QY 120 DMLTJHIGSSPEKPYLYLVKVGSKQETAKNAIMIDCGIHAREMISPACLMFI----- 173
DB 138 EWSRKIGISTVEDNPLVYKI-GEKNERKALFMDCGIHAREMWSPAFCQWVYQATKT 196
QY 174 -----GH-----NRMRKRNSFYANNHCIGTDLNNRFAS 202
DB 197 YGRNKIMTKLLDMNFIYILFVFNVDGYIMSWTRNMRKRNSKONSKCIGTDLNNRF-N 255
QY 203 KHWCEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYQHIYPPYS 262
DB 256 ASWNSIPNTNDPCADNIRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYQMLFPPYG 315
QY 263 YTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHSFETLYLAPGGDDMIYDLGIKYS 322
DB 316 YTSKLPNNHEDLAKVAKIGTDVL-STREYETRYIYGPIESTIYISGSSLDMAWDLGIKHT 374
QY 323 F 323
DB 375 F 375

```

```

RESULT 12
US-08-430-787A-6
: Sequence 6, Application US/08430787A
: Patent No. 5593674
:
: GENERAL INFORMATION:
: APPLICANT: Drayna, Dennis T., Eaton, Dan L.
: TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patln (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/430,787A
: FILING DATE: 27-APR-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/277,540
: FILING DATE: 19-JUL-1994
: APPLICATION NUMBER: 08/167727
: FILING DATE: 15-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/959944
: FILING DATE: 14-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/649591
: FILING DATE: 01-FEB-91
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616

```

```

? REFERENCE/DOCKET NUMBER: 669D9D1D1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-1896
? TELEFAX: 415/952-9881
? TELE: 910/371-7168
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 417 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
US-08-430-787A-6

```

Query Match	34.68;	Score 622;	DB 1;	Length 417;
Best Local Similarity	35.58;	Pred. No. 1.2e-58;		
Matches 128;	Conservative 71;	Mismatches 118;	Indels 44;	Gaps 7

[illegible]

RESULT 13
US-09-171-945-125
; Sequence 125, Application US/09171945
; Patent No. 6377500

APPLICANT: Emery, Stephen
 APPLICANT: Copley, Clive Graham
 APPLICANT: Edge, Michael Derek
 TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said
 TITLE OF INVENTION: Antibody, and Their Therapeutic Use in an Adept System
 FILE REFERENCE: Monoclonal Antibody to CEA
 CURRENT APPLICATION NUMBER: US/09/171,945
 PRIOR FILING DATE: 1998-10-29
 PRIOR APPLICATION NUMBER: GB9703103.3
 PRIOR FILING DATE: 1997-02-14
 PRIOR APPLICATION NUMBER: GB9609405.7
 PRIOR FILING DATE: 1996-05-04
 PRIOR APPLICATION NUMBER: PCT/GB97/01165
 PRIOR FILING DATE: 1997-04-29
 NUMBER OF SEQ ID NOS: 131
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO: 125
 LENGTH: 716
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: humanized

US-09-171-945-125

Query Match	34.5%;	Score 619.5;	DB 4;	Length 716;
Best Local Similarity	35.2%;	Pred. No. 5.5e-58;		
Matches 127;	Conservative	69;	Mismatches 122;	Indels 43;
			Gaps	6

[illegible]

RESULT 14
US-08-782-760-6
; Sequence 6, Application US/08782760
; Patent No. 5948668
; GENERAL INFORMATION:

APPLICANT: Hartman, Jacob
APPLICANT: Fulga, Netta
APPLICANT: Mendelovitch, Simona
APPLICANT: Gorecki, Marian
TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
TITLE OF INVENTION: CARBOXYPEPTIDASE B
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1 0, Version #1 30
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 1154097383 7560

```

CLASSIFICATION: 43F
FILING DATE: 13-JAN-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/378,233
FILING DATE: 25-JAN-1995

ATTORNEY/AGENT INFORMATION:
NAME: White, John P

REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 62

ADVICE NUMBER: 0336/4384/
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:45:11 ; Search time 65 Seconds

(without alignments)
577.584 Million cell updates/sec

Title: us-09-980-881-3

Perfect score: 1798
Sequence: 1 FOSQVLAALPRISROYVLT.....IKTSFTSNPEVKLLPLSLK 338

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

A.Geneseq_032802:*
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1798	100.0	338	21	AAB11458 Human brain carbox
2	1785	99.3	360	21	AAB11457 Human brain carbox
3	1699.5	94.5	423	17	AAR90293 Human plasma carbo
4	1699.5	94.5	423	18	AAW14733 Human plasma carbo
5	1699.5	94.5	423	20	AAW92270 Human plasma carbo
6	1694.5	94.2	423	14	AAR36273 Human plasma carbo
7	1338	74.4	246	21	AAB11459 Human brain carbox
8	816.5	45.4	211	21	AAB58129 Lung cancer associ
9	646.5	36.0	404	16	AAR75131 Porcine Tyr-His-Me
10	637.5	35.5	415	17	AAW06172 Human pancreatic c
11	637.5	35.5	415	19	AAW74476 Human pancreatic c

12	637.5	35.5	424	17	AAW06175
13	636.5	35.4	324	22	AA665562
14	619.5	34.5	716	18	AAW41414
15	612	34.0	437	22	AAE00502
16	610	33.9	307	17	AAW00602
17	582	32.4	306	16	AAW75132
18	582	32.4	306	22	AAU04477
19	569	31.6	354	22	AA665560
20	569	31.6	354	22	AA665566
21	569	31.6	354	22	AA665566
22	569	31.6	374	22	AA665547
23	569	31.6	374	22	AA665547
24	567	31.5	329	17	AAW06174
25	567	31.5	349	17	AAW06173
26	560	31.1	349	17	AAW06181
27	560	31.1	349	17	AAW06181
28	559	31.1	349	17	AAW06182
29	559	31.1	349	17	AAW06182
30	556	30.8	349	18	AAW13750
31	554	30.8	349	18	AAW13753
32	554	30.8	349	18	AAW13752
33	554	30.8	349	22	AAE01664
34	553	30.8	349	18	AAW13751
35	552	30.7	349	18	AAW13753
36	551	30.6	311	22	AA665563
37	551	30.6	349	18	AAW13760
38	551	30.6	349	18	AAW13761
39	551	30.6	349	18	AAW13762
40	549	30.5	349	18	AAW13757
41	549	30.5	349	18	AAW13765
42	549	30.5	613	19	AAW41412
43	548	30.5	349	18	AAW74481
44	548	30.5	349	18	AAW13759
45	546	30.4	349	18	AAW13764

ALIGNMENTS

RESULT 1	
ID	AAB11458 standard; Protein; 338 AA.
XX	
AC	AAB11458;
XX	
DT	01-MAR-2001 (first entry)
XX	
DE	Human brain carboxypeptidase B protein SEQ ID NO 3.
XX	
KW	Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW	treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW	cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW	hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW	Down's syndrome; head trauma.
XX	
OS	Homo sapiens.
XX	
PN	WO200066717-A1.
XX	
PD	09-NOV-2000.
XX	
PF	01-MAY-2000; 2000WO-JP02878.
XX	
PR	30-APR-1999; 99JP-0125169.
XX	
PA	(MATS/) MATSUMOTO A.
XX	
PI	Matsumoto A;
XX	
DR	WPI; 2000-687534/67.
XX	
PT	Human brain carboxypeptidase B isolated from the hippocampus useful for screening agents for the treatment of Alzheimer's and other brain

PT disorders -
 XX
 PS Disclosure; Page 71-73; 84pp. Japanese.
 XX
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity, and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 XX
 SO Sequence 338 AA;

Query Match 100.0%; Score 1798; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 1, 2e-179;
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQSGVLAALPRTSRQVYQVONLTITTEYIVLMQPVADLIYKKQVHFVNASDVNVKA 60
 DB 1 fgsqgvlaalprrtsrqvylqnlitlyelvlwqvtcadllvkkqvhffvnasdvndvka 60
 QY 61 HLNVSGIPGSVLLADVEDLIQOOISNDTVSPRASASYEYOHSHNETYSMTETTERHPD 120
 DB 61 hlrvsglpcsvlladvedllqgqisndtvsprrasasyeqyhslnelyswlelterhpd 120
 QY 121 MLTKIHGSSFEKYPPLYLVKVGSKQETAKNAIWDGCIHAREWISPAFCIMFIGNRMWR 180
 DB 121 mltkihgssfekyplylvkvgskqetaknaikwdcgiharewispafclwfighnmwr 180
 QY 181 KNRSEYANNHCIGTDLRNPFASKHMCCEGASSSCSETYCGLPSEPEYKAVASFLRRN 240
 DB 181 knrsfyanhncigtldlrnrfaskhmcceegassscsetyctylpsepevkavasflrrn 240
 QY 241 INQIKAYISMHSYQHIIVFPYSYTRSKSKDHEELSLVAEAVRAIEKTSKNTRYTHGHS 300
 DB 241 inqikayismhsyqhiivfpystrskskdheelslvaseavraidektskntrythngs 300
 QY 301 ETLYLAPGGGDWYIDGIKYFTSNPPVEKLLPLSLK 338
 DB 301 etlylapgggdwlydglkyfstsnpvекllpislk 338

RESULT 2
 AAB11457
 ID AAB11457 standard; Protein: 360 AA.
 XX
 AC AAB11457;
 XX
 DT 01-MAR-2001 (first entry)
 XX
 DE Human brain carboxypeptidase B protein.
 XX
 KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.
 XX
 OS Homo sapiens.
 XX
 XX WO200006717-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 01-MAY-2000; 2000MO-JP02878.
 XX
 PR 30-APR-1999; 99JP-0125169.
 XX
 PA (MATS/) MATSUMOTO A.

XX
 PI Matsumoto A;
 XX
 DR WPI: 2000-687534/67.
 DR N-PSDB: AAC81962.
 XX
 PT Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX
 PS Claim 1; Page 68-71; 84pp; Japanese.
 XX
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity, and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 XX
 SO Sequence 360 AA;

Query Match 99.3%; Score 1785; DB 21; Length 360;
 Best Local Similarity 99.1%; Pred. No. 3e-178;
 Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FQSGVLAALPRTSRQVYQVONLTITTEYIVLMQPVADLIYKKQVHFVNASDVNVKA 60
 DB 23 fgsqgvlaalprrtsrqvylqnlitlyelvlwqvtcadllvkkqvhffvnasdvndvka 82
 QY 61 HLNVSGIPGSVLLADVEDLIQOOISNDTVSPRASASYEYOHSHNETYSMTETTERHPD 120
 DB 83 hlrvsglpcsvlladvedllqgqisndtvsprrasasyeqyhslnelyswlelterhpd 142
 QY 121 MLTKIHGSSFEKYPPLYLVKVGSKQETAKNAIWDGCIHAREWISPAFCIMFIGNRMWR 180
 DB 143 mltkihgssfekyplylvkvgskqetaknaikwdcgiharewispafclwfighnmwr 202
 QY 181 KNRSEYANNHCIGTDLRNPFASKHMCCEGASSSCSETYCGLPSEPEYKAVASFLRRN 240
 DB 203 knrsfyanhncigtldlrnrfaskhmcceegassscsetyctylpsepevkavasflrrn 262
 QY 241 INQIKAYISMHSYQHIIVFPYSYTRSKSKDHEELSLVAEAVRAIEKTSKNTRYTHGHS 300
 DB 263 inqikayismhsyqhiivfpystrskskdheelslvaseavraidektskntrythngs 322
 QY 301 ETLYLAPGGGDWYIDGIKYFTSNPPVEKLLPLSLK 338
 DB 323 etlylapgggdwlydglkyfstsnpvекllpislk 360

RESULT 3
 AAR90293
 ID AAR90293 standard; Protein: 423 AA.
 XX
 AC AAR90293;
 XX
 DT 12-APR-1996 (first entry)
 XX
 DE Human plasma carboxypeptidase B.
 XX
 KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
 KW purification; plasminogen; affinity column.
 XX
 OS Homo sapiens.
 XX
 XX
 XX
 XX
 FT Key Location/Qualifiers
 FT Active-site 181
 FT /note= "Residue suspecting of being involved in
 FT catalytic activity of enzyme."

FT	Binding-site	183	/note=	"Residue suspected of being part of the binding site of enzyme."
FT	Binding-site	256..257	/note=	"Residues suspected of being part of the binding site of enzyme."
FT	Active-site	310	/note=	"Residue suspected of being involved in catalytic activity of enzyme."
FT	Binding-site	312	/note=	"Residue suspected of being part of the binding site of enzyme."
FT	Active-site	369	/note=	"Residue suspected of being involved in catalytic activity of enzyme."
FT	Blinding-site	370..371	/note=	"Residues suspected of being part of the binding site of enzyme."
FT	Blinding-site	394	/note=	"Residue suspected of being part of the binding site of enzyme."
XX	US5474901-A.			
PD	12-DEC-1995.			
XX	01-FEB-1991;	91US-0649591.		
PR	01-FEB-1991;	91US-0649591.		
PR	14-OCT-1992;	92US-0959944.		
PR	15-DEC-1993;	93US-0167727.		
PR	19-JUL-1994;	94US-0277540.		
PA	(GETH) GENENTECH INC.			
PI	Drayna DT, Eaton DL:			
DR	WPI; 1996-039508/04.			
DR	N-PSDB; AAT11671.			
PT	Antibody to human plasma carboxypeptidase B - useful for detecting and purifying hpcpb for use in treating clotting disorders e.g. haemophilia A			
CC	Disclosure: Figure 4; 40pp; English.			
CC	An antibody which specifically binds human plasma carboxypeptidase B (hpcpb) and does not cross react with other carboxypeptidases is useful for the detection of hpcpb in vitro. The antibody is also used for purifying hpcpb from a sample. Purification comprises passing a sample thought to contain hpcpb over either a column to which antibody has been bound, or a plasminogen affinity column, eluting the column and then recovering the fraction containing the hpcpb.			
SQ	Sequence 423 AA:			
Query Match	94.5%	Score 1699.5;	DB 17;	Length 423;
Best Local Similarity	89.8%;	Pred No. 3.6e-169;		
Matches 324; Conservative	0;	Mismatches 0;	Indels 37;	Gaps 1
OY	1 FOSGOVLAALPPTSROVOVLONLITTYEIVELMQPWYADLIVRKQVHFVNASDVNPKA	60		
Dd	23 fgsqgvalaalprtsrqvqlgnlttelyelivqpvradilvkkyghfftnasadvnvka	82		
OY	61 HLNVSIGFPCSVLLADVEDLIQQOISNDTVSPRASAIFYEQYHSLNEIYSWIEFTERHPD	120		
Dd	83 hlnvsigfpcsvlladvedliqqisndtvspasasyeqymslneiywiefterhp d	142		
OY	121 MLTKHGSSFEKYLYLVKVGSGKEGTAKNALIMIDCGIHAREMISAPFLMFIGH-----	175		
Dd	143 mltklhgsstfekylylvkvgsgkegtaknalwldcglharewispaflwifighltqfy	202		

QY	176	-----NMWKKNSFEVANNICIGDILNRPNASK	203
Db	203	g1lqytnlll1ldvfymprvvrvndgydyswkknmmwtknsfYannncigtldlnrn1ask	262
QY	204	HMCEGASSSCSETCGLYPESEPEVKAVASFLRRNIINIKAVISMHYSQHIIVPEYSY	263
Db	263	hmcegaassscsetcglYPEsepevkavasflrrnlnqkaylsmhsysqhlvfpyf	322
QY	264	TRRSKDHHEELSVASENVRAIENTSKNTRTHHGETLYTLAPGGDDWIYDLGRTYSF	323
Db	323	trrskdhheel1svaseavra1ektskntrythghsetlylapggddw1ydlgkysf	382
QY	324	T	324
Db	383	t	383
RESULT 4			
ID	AAW14733	AAW14733 standard; Protein: 423 AA.	
XX	AAW14733;		
AC			
XX	08-MAY-1997	(first entry)	
DT			
XX			
DE		Human plasma carboxypeptidase B.	
XX			
KW		Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;	
KW		plasma; plasminogen.	
XX			
OS		Homo sapiens.	
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..22	
FT		/note- "Signal peptide"	
FT	Protein	23..423	
FT		/note- "Mature PCPB"	
FT	Cleavage-site	114	
FT		/note- "Potential clip site for activation of PCPB as a carboxypeptidase"	
FT	Active-site	181	
FT		/note- "Expected to be involved in catalytic activity	
FT	Binding-site	183	
FT		/note- "Expected to be involved in substrate binding"	
FT	Active-site	184	
FT		/note- "Expected to be involved in catalytic activity	
FT	Binding-site	256	
FT		/note- "Expected to be involved in substrate binding"	
FT	Binding-site	257	
FT		/note- "Expected to be involved in substrate binding"	
FT	Active-site	310	
FT		/note- "Expected to be involved in catalytic activity	
FT	Binding-site	312	
FT		/note- "Expected to be involved in substrate binding"	
FT	Active-site	369	
FT		/note- "Expected to be involved in catalytic activity	
FT	Binding-site	370	
FT		/note- "Expected to be involved in substrate binding, thought to determine specificity of PCPB as a carboxypeptidase B"	
FT	Binding-site	371	
FT		/note- "Expected to be involved in substrate binding"	
FT	Binding-site	394	
FT		/note- "Expected to be involved in substrate binding"	
XX	US5593674-A.		
PN			
XX	14-JAN-1997.		
PD			
XX			
PR	01-FEB-1991;	91US-0649591.	
XX			
PR	01-FEB-1991;	91US-0649591.	
XX			

PR 14-OCT-1992; 92US-0959944.
 PR 15-DEC-1993; 93US-0167727.
 PR 19-JUL-1994; 94US-0277540.
 PR 27-APR-1995; 95US-0430787.

PA (GETH) GENENTECH INC.

PI Drayna DT, Eaton DL;

DR WPI; 1997-099413/09.

DR N-PSDB; AAT62846.

PT Using human plasma carboxypeptidase B in blood coagulation - is
 PT functionally related to carboxypeptidase A and pancreas
 PT carboxypeptidase B

PS Example 2; Column 37-42; 39pp; English.

CC This sequence represents human plasma carboxypeptidase B (PCPB) which
 CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
 CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
 CC purified from human plasma or by transformed cell culture by
 CC extraction using plasminogen bound to a solid phase.

SQ Sequence 423 AA;

Query Match 94.5%; Score 1699.5; DB 18; Length 423;
 Best Local Similarity 89.8%; Pred. No. 3,6e-169;

Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 FOSGQVLAALPRTSRQVOVLQNTTTEYELVMOPTADLIYKKQYHFFVNASDVNDYKA 60
 DB 23 fgsqgvlaalprtsrqvqlnlttelyelwqvtadliivkkqyhfivnasdvndyka 82
 QY 61 HUNSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSINEIYSWIEFITERHPD 120
 DB 83 hlnvsgipcsvlladvedliqqisndtvsprrasasyeqyhsineiyswiefterhpd 142
 QY 121 MLTKIHGSSFEKPYLYLVKSGKEOTAKNAWIDCGIHARWISPAFLMFTIGH----- 175
 DB 143 mltkihgssfekpylylvksgkeotaknalwldcgiharewispafclwflghtqfy 202
 QY 176 -----NRMKRKNRSPYANNHCIGTDLNRFFASK 203
 DB 203 gllgqytnllrlvdfymvpmvndgydyswkknmwrknrsfyannhcgldlnrnfsak 262
 QY 204 HMCEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNINQIKAYISMSHSOHIFFPSY 263
 DB 263 hwceegasssscseycgylpesepevkavasflrrninqikayismhsyqhlvffpsy 322
 QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLVAPGGGDWYIDGIRKYSF 323
 DB 323 trsksdheelslvaseavraiektsknctrythngseetlvapgggdwlydigrkysf 382
 QY 324 T 324
 DB 383 t 383

RESULT 5

AAW92270 ID AAW92270 standard; Protein; 423 AA.

XX AAW92270;

XX 28-APR-1999 (first entry)

DE Human plasma carboxypeptidase B (PCPB) thr147.

KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
 XX polymorphism detection; thrombotic disease.

OS Homo sapiens.

EH Key Location/Qualifiers

FT Peptide 1..22

FT Protein /note="signal peptide"

FT /note="mature PCPBthr147"

PN WO9855645-A1.

PD 10-DEC-1998.

XX 02-JUN-1998; 98MO-EP03244.

PR 03-JUN-1997; 97US-0869057.

XX (SCHD) SCHERING AG.

PI Morser MJ, Nagashima M;

DR WPI; 1999-045800/04.

DR N-PSDB; AAV74302.

PT Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 PT humans

PS Disclosure: Page 25-26; 35pp; English.

CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
 CC hPCPBthr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphisms of the
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.

XX Sequence 423 AA;

Query Match 94.5%; Score 1699.5; DB 20; Length 423;
 Best Local Similarity 89.8%; Pred. No. 3,6e-169;

Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 FOSGQVLAALPRTSRQVOVLQNTTTEYELVMOPTADLIYKKQYHFFVNASDVNDYKA 60
 DB 23 fgsqgvlaalprtsrqvqlnlttelyelwqvtadliivkkqyhfivnasdvndyka 82
 QY 61 HUNSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEQYHSINEIYSWIEFITERHPD 120
 DB 83 hlnvsgipcsvlladvedliqqisndtvsprrasasyeqyhsineiyswiefterhpd 142
 QY 121 MLTKIHGSSFEKPYLYLVKSGKEOTAKNAWIDCGIHARWISPAFLMFTIGH----- 175
 DB 143 mltkihgssfekpylylvksgkeotaknalwldcgiharewispafclwflghtqfy 202
 QY 176 -----NRMKRKNRSPYANNHCIGTDLNRFFASK 203
 DB 203 gllgqytnllrlvdfymvpmvndgydyswkknmwrknrsfyannhcgldlnrnfsak 262
 QY 204 HMCEGASSSSCSEYTCGLYPESEPEVKAVASFLRRNINQIKAYISMSHSOHIFFPSY 263
 DB 263 hwceegasssscseycgylpesepevkavasflrrninqikayismhsyqhlvffpsy 322
 QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLVAPGGGDWYIDGIRKYSF 323
 DB 323 trsksdheelslvaseavraiektsknctrythngseetlvapgggdwlydigrkysf 382

Db	323	trkskdhnealslvaseavraiektsntrylthngsetlylapgggddwlydlgkysf	362
Oy	324	T 324	
Db	383	t 383	
RESULT 6			
XX	AAAR36273	AAAR36273 standard; Protein; 423 AA.	
AC	AAAR36273;		
XX	24-AUG-1993	(first entry)	
XX	Human plasma carboxypeptidase B.		
DE	PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.		
XX	Homo sapiens.		
XX	Key	Location/Qualifiers	
FT	Peptide	1..22	
FT	Cleavage-site	114	
FT	/note=	"putative site of cleavage by trypsin to activate PCPB as a carboxypeptidase"	
FT	Active-site	181	
FT	/note=	"expected to be involved in catalytic activity"	
FT	Binding-site	183	
FT	/note=	"expected to be involved in substrate binding"	
FT	Active-site	184	
FT	/note=	"expected to be involved in catalytic activity"	
FT	Binding-site	256..257	
FT	/note=	"expected to be involved in substrate binding"	
FT	Active-site	310	
FT	/note=	"expected to be involved in catalytic activity"	
FT	Binding-site	312	
FT	/note=	"expected to be involved in substrate binding"	
FT	Active-site	369	
FT	/note=	"expected to be involved in catalytic activity"	
FT	Binding-site	370..371	
FT	/note=	"expected to be involved in substrate binding; Asp 370 is the residue which determines specificity of the PCPB as a carboxypeptidase B"	
FT	Binding-site	394	
FT	/note=	"expected to be involved in substrate binding"	
XX	US5206161-A.		
XX	27-APR-1993.		
XX	01-FEB-1991;	91US-0649591.	
XX	01-FEB-1991;	91US-0649591.	
XX	(GENTH) GENENTECH INC.		
XX	Drayna DT, Eaton DL;		
XX	WPI; 1993-151724/18.		
XX	N-PSDB; AAQ41001.		
XX	New human plasma carboxypeptidase B - used as haemostatic		

[illegible]

XX 09-NOV-2000.
 PD
 XX
 XX 01-MAY-2000; 2000MO-JP02878.
 PF
 XX
 XX 30-APR-1999; 99JP-0125169.
 PR
 XX
 XX (MATS/) MATSUMOTO A.
 PA
 XX
 XX MATSUMOTO A;
 PI
 XX
 XX WPI; 2000-687534/67.
 DR
 XX
 XX Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 PS
 XX Disclosure; Page 74-75; 84pp; Japanese.
 CC
 XX This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimers,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 XX Sequence 246 AA;
 SQ
 Query Match 74.4%; Score 1338; DB 21; Length 246;
 Best Local Similarity 100.0%; Pred. No. 1.2e-131;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 93 ASASYEQHSHNTEYSWIEFTEHRHMDLTKIHGSSFEKYLVLKSGKEQTAKNAI 152
 DB 1 aasasyeqhshnelyswieftehrhmdltnhlgssfekylvlksgkeqtaknai 60
 OY 133 WIDCGIHAREWISPAFCIMFIGNHMKRKNRSFYANNHCIGTDLNRNFASKHWCCEGASS 212
 DB 61 widcgiharewispaflcwmfignhmkrsfyannhncigtdlnrnflaskhwccegass 120
 OY 213 SCSSEYCGIYPESPEPKAVASFLRNINOKATISMHSYSGHIVFYSTRSKDHE 272
 DB 121 sscsetcgylypesepekavassflrlninqikayismhsyqhlvfysylrskskdhe 180
 OY 273 ELISIVASEAVRAIEKTSKNTRTYTHGSEETLYLAPGGDDWYDGIKYSFTSNPVEKL 332
 DB 181 elisvaseavraiektskntrtythgsetcylapggddwlydglkysftsnpvekl 240
 OY 333 LPLSLIK 338
 DB 241 lplslk 246
 RESULT 8
 AAB58129
 ID AAB58129 standard; Protein: 211 AA.
 XX
 AC AAB58129;
 XX
 XX 14-MAR-2001 (first entry)
 DT
 XX
 DE Lung cancer associated polypeptide sequence SEQ ID 467.
 XX
 XX Human; lung cancer associated protein; neuroprotective; cyostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnary;
 KW gastrointestinal; nephrotropic; antinfecive; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease;
 XX

OS Homo sapiens.
 XX
 XX WO200055180-A2.
 PN
 XX
 XX 21-SEP-2000.
 PD
 XX
 XX 08-MAR-2000; 2000MO-US05918.
 PF
 XX
 XX 12-MAR-1999; 99US-0124270.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX (ROSE/) ROSEN C A.
 PI
 XX
 XX Ruben SM;
 DR
 XX WPI; 2000-587514/75.
 DR N-PSDB; AAF18005.
 XX
 XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 PS
 XX Claim 11; Page 944-945; 1425pp; English.
 CC
 XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cyostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnary; gastrointestinal
 CC general; nephrotropic; antinfecive; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as research purposes, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 SQ Sequence 211 AA;
 Query Match 45.4%; Score 816.5; DB 21; Length 211;
 Best Local Similarity 90.6%; Pred. No. 4.7e-77;
 Matches 154; Conservative 1; Mismatches 6; Indels 9; Gaps 2;
 OY 163 WISPAFCIMFL-----CHNRMKRKNRSFYANNHCIGTDLNRNFASKHWCCEGASS 214
 DB 3 wis-mcrlwmymymyswknrmwkrnrfyannhncigtdlnrnflaskhwccegass 61
 OY 215 CSETYCGIYPESPEPKAVASFLRNINOKATISMHSYSGHIVFYSTRSKDHEEL 274
 DB 62 csetcgylypesepekavassflrlninqikayismhsyqhlvfysylrskskdhe 121
 OY 275 SLVASEAVRAIEKTSKNTRTYTHGSEETLYLAPGGDDWYDGIKYSFT 324
 DB 122 slvaseavraiektskntrtythgsetcylapggddwlydglkysft 171
 RESULT 9
 AAR75131
 ID AAR75131 standard; Protein: 404 AA.
 XX
 AC AAR75131;
 XX
 XX 13-MAR-1996 (first entry)
 DT
 XX
 DE Porcine Tyr-His-Met Procarboxypeptidase B.
 XX


```

PR 16-AUG-1995: 95GB-0016810.
PR 23-DEC-1994: 94GB-0026192.
XX
XX (ZENE ) ZENECA LTD.
PA
PI Blakey DC, Boyle FT, Davies DH, Eggelte HT, Heaton DW,
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
XX WPI; 1996-321650/32.
DR N-PSDB; AAT42506.
XX
XX Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
PT anti-neoplastic prodng
XX
XX Reference Example 18; Page 140-142; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC prodng. The system is based on antibody directed enzyme prodng therapy
CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC be used. The present sequence is that of proHCPB with a PelB leader
CC sequence encoded by PIC11738.
XX
XX Sequence 424 AA:
SQ
Query Match 35.5%; Score 637.5; DB 17; Length 424;
Best Local Similarity 36.0%; Pred. No. 8,6e-58;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;
OY 1 FSGGVIALPRTSRQVYQNIITTYEIVLMQPVYADLIVKKQVHFPVNASDVNVA 60
DB 29 fgekvfrvrvdenhnlirelastqldfwkpsdvqtkphstvdfrvkaedltven 88
OY 61 HUNVSGIPSCVLLADVEDLIQOOISNDTVSPRASASYEYOYHSLNDIYSWIEFTRHPD 120
DB 89 vlkqnelgykvvlslmnlrvveagfdrv---ratgshyekynkhwetlsgvtenpa 145
OY 121 MTKIHIGSFEEKYPLYLVKVGSEKQTAKNAIWDGTHAREWISPAFLMFI----- 173
DB 146 llsrvslgtfegralylkv-gkagqnpkafimdcgfharwlsfpafqvfreavrtv 204
OY 174 -----GH-----NRMWRKNSRFYANNHCITGDLNRNFRASK 203
DB 205 greiqvteflldkldfylvplvlnldgylytwktsrftwkrktsthtgsscldgdpnrnf-da 263
OY 204 HMCERGAASSSCSEYVGLYPSEPEVKAVASFLRRNINQIKAYISMHSYQHIVPFYSY 263
DB 264 gwceigasnpecdetcycaaseketkaladflrnklsikyllhaysqmmilypsey 323
OY 264 TMSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSEFLYLAPGGGDWITDLGATKYSF 323
DB 324 aykligenaealnakatvkel-aslhgtkytgyppattlpyaagsddwaydqgltysf 382
OY 324 T 324
DB 383 t 383
XX
XX RESULT 13
XX AAG6562
XX AAG6562 standard; Protein; 324 AA.
XX
XX AAG6562;
XX
XX 22-OCT-2001 (first entry)
XX
XX Bothops jararaca carboxypeptidase homologue protein.
XX

```

```

XX secreted metallocarboxypeptidase; immunomodulatory;
XX antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
XX osteopathic; antiarthritic; cerebroprotective; vasotropic; vulnerary;
XX antibacterial; antiviral; antifungal; antiinflammatory; antidiabetic;
XX antineumatic; cytostatic; antiallergic; haemostatic; infection;
XX cancer.
XX
XX Bothops jararaca.
XX
XX WO200157265-A1.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-0503783.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 29-SEP-2000; 2000US-0676135.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Boyle BJ, Mize NK, Arterburn MC, Yeung G, Tang YT, Lau C;
XX Drmanac RT, Wang M, Chen L, Yang Y;
XX WPI; 2001-465705/50.
XX
XX Novel metallocarboxy peptidase-like polypeptide is useful in treating
XX neurodegenerative, immune, etc. disorders
XX
XX Example 3; Fig 1; 146pp; English.
XX
XX The invention relates to an isolated polypeptide comprising an amino
XX acid sequence of 374, 41, 14, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
XX 354, 165, 374, 354, or 41 amino acids, fully defined in the
XX specification, or its translated protein coding portion, mature protein
XX coding portion, extracellular portion or active domain. The polypeptide
XX is a human secreted metallocarboxypeptidase-like polypeptide. It
XX is useful for treating mammalian diseases including:
XX neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
XX osteoporosis, osteoarthritis, disorders linked to reduced tissue
XX growth, Huntington's disease, amyotrophic lateral sclerosis, stroke,
XX reperfusion, wound healing, infectious diseases, immune and autoimmune
XX disorders such as severe combined immunodeficiency, rheumatoid
XX arthritis, guillain-Barre syndrome, insulin dependent diabetes
XX mellitus, graft-versus-host disease and allergy, cancer, and
XX haemophilia. The polypeptide is also useful in medical imaging, as an
XX immunogen to generate antibodies, as a nutritional source or
XX supplement, and as a contraceptive. The present sequence is
XX a carboxypeptidase homologue protein of Bothops jararaca.
XX
XX Sequence 324 AA:
SQ
Query Match 35.4%; Score 636.5; DB 22; Length 324;
Best Local Similarity 42.0%; Pred. No. 7.1e-58;
Matches 126; Conservative 54; Mismatches 77; Indels 43; Gaps 5;
OY 62 LNVSGIPSCVLLADVEDLIQOOISNDTVSPRASASYEYOYHSLNDIYSWIEFTRHPDM 121
DB 2 lqsgslnyellnldgvlrdqldnha---rtagynyekynskewkladwaddanlanspl 58
OY 122 LTRIHIGSFEEKYPLYLVKVGSEKQTAKNAIWDGTHAREWISPAFLMFI----- 173
DB 59 vsrlqigtftgfpmpilkv-gkpyvnhkalfidcgfharewlsfpafqvfreavrtv 117
OY 174 -----GHNMWRKNSRFYANNHCITGDLNRNFRASKH 204
DB 118 ketimqljlnklldfylvplvlnldgyvywkwqsrwmwktstsvnagstcldgdpnrnfdaa- 176
OY 205 WCEGASASSSCSEYVGLYPSEPEVKAVASFLRRNINQIKAYISMHSYQHIVPFYSY 264
DB 177 wcsvgarnpcsetygskspeketkaladflrnsltdqytlthysqmmilypsytl 236
XX

```

QY 265 RSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWYDLGIRYSFT 324
 Db 237 ydltsnknklnslakealreel-kvligteyly9p9aaltlypaagsgddwaydgqilyatf 295

RESULT 14
 AAM41414
 ID AAM41414 standard; Protein: 716 AA.

XX AAM41414;

DT 02-JUN-1998 (first entry)

DE PreproHCPB-linker-humanised 806.077 VH.

KM Anti-CEA antibody; carcinoembryonic antigen; 806.077 Ab; cancer therapy;
 KM cancer diagnosis; complementarity determining region.

OS Synthetic.

PN WO9742329-A1.

PD 13-NOV-1997.

PF 29-APR-1997; 97WO-GB01165.

PR 14-FEB-1997; 97GB-0003103.
 PR 04-MAY-1996; 96GB-0009405.

PA (ZENEC) ZENECAL LTD.

PI Copley CG, Edge MD, Emery SC;

DR WPI: 1997-558987/51.

DR N-PSDB: AAV17331.

PT Anti-carcinoembryonic antigen antibody 806.077 Ab - used for
 PT diagnosis and therapy of cancer

PS Example 75; Page 184-187; 208pp; English.

XX This sequence is preproHCPB-linker-humanised 807.077 VH, which is
 CC an example of the antibody of the invention. The antibody is an anti-CEA
 CC (carcinoembryonic antigen) antibody (preferably 806.077 Ab). Host cells
 CC or transgenic organisms transformed with DNA encoding the antibody, are
 CC used to make the antibody or conjugate. The conjugate is used in a
 CC medicament suitable for intravenous administration. The conjugate can be
 CC used for cancer therapy, selectively killing tumour cells. The antibody
 CC can be used for in vivo or in vitro diagnosis of cancer.

SQ Sequence 716 AA;

Query Match 34.5%; Score 619.5; DB 18; Length 716;
 Best Local Similarity 35.2%; Pred. No. 1.6e-55;
 Matches 127; Conservative 69; Mismatches 122; Indels 43; Gaps 6;

QY 1 FOSGOVLALPRTSRQOVYONLTTTEYIVLMQPTADLYKKQVFFPNASVDNVKA 60

Db 22 fegekvtrvnedenhnlitrelastqldfkwkpsvltqkphstvfrykaedtlven 81

QY 61 HLANVSGIPCSVLADVDLIIQOQISNDTVSPRASASYEQQVHSLNETYSMTIEFTTERHPD 120

Db 82 vlkqnelqyvlvllsnlnrveaqdldarv----ratghseyekynkwtleatvtqyatenpa 138

QY 121 MLRKIHGSSFEKPYLVLVKSGKEQTAKNAIWDICGIIHAREWISPAFLMFI----- 173

Db 139 llsrsvlgttfefgralylklv-gkaggnkpaifmdcgfharewlspsfcqgvtvreatvty 197

QY 174 -----GH-----NRMRKNSFANNHCITDILNRNASK 203

Db 198 greldqtelldkldfyvlpvlnldgylvtwcksrftwcktrsthtsgscldtqdpnrf-da 256

QY 204 HMCDEGASSSCSEYTCGLYPESEPEVKAVASFLRRNINOIKAYISMHSQHIYFYSY 263
 Db 257 gwceigaasnpdcelycgaaseketakaladflrnklsstlksayltltnsqmalyrpy 316

QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWYDLGIRYSF 323

Db 317 ayklgenaelnalakatvkel-aslhgtklytgygacttlypsagtsdwdydgqlyrsf 375

QY 324 T 324

Db 376 t 376

RESULT 15
 AAE00502
 ID AAE00502 standard; Protein: 437 AA.

XX AAE00502;

DT 19-JUN-2001 (first entry)

DE Human carboxypeptidase-related protein #5.

KM Human; carboxypeptidase-related protein; protease; gene therapy;
 KM drug screening; mental; biological disorder; medical disorder.

OS Homo sapiens.

PN WO200123568-A1.

PD 05-APR-2001.

PF 29-SEP-2000; 2000WO-US26876.

PR 29-SEP-1999; 99US-0156685.

PA (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

DR WPI: 2001-266171/27.

DR N-PSDB: AAD03837.

PT New isolated human carboxypeptidase polynucleotide useful for
 PT generating antibodies, as reagents in diagnostic assays and for
 PT screening for compounds useful for treating mental, biological or
 PT medical diseases

PS Claim 2; Page 30-31; 36pp; English.

XX The present sequence is a novel human protein (NHP),
 CC carboxypeptidase-related protein. The carboxypeptidase-related
 CC proteins share structural similarity with animal proteases, and
 CC especially carboxypeptidase B or carboxypeptidase A. The
 CC carboxypeptidase-related protein cDNA is useful for the detection of
 CC mutant human carboxypeptidases, or inappropriately expressed human
 CC carboxypeptidases for the diagnosis of disease, and also as a
 CC therapeutic. It is also useful in drug screening, for generation of
 CC antibodies, as reagents in diagnostic assays, for the identification of
 CC other cellular gene products related to human carboxypeptidases, and as
 CC reagents in assays for screening for compounds that are useful for
 CC treating mental, biological or medical disorders and diseases. Nucleotide
 CC constructs encoding functional NHPs can also be used in gene therapy.

SQ Sequence 437 AA;

Query Match 34.0%; Score 612; DB 22; Length 437;
 Best Local Similarity 35.4%; Pred. No. 4.3e-55;
 Matches 129; Conservative 62; Mismatches 127; Indels 46; Gaps 7;

QY 1 FOSGOVLALPRTSRQOVYONLTTTEYIVLMQPTADLYV-KQVHFFPNASVDNV 58

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:49:04 ; Search time 55.22 Seconds
(without alignments)
1058.897 Million cell updates/sec

Title: US-09-980-881-3

Perfect score: 1798
Sequence: 1 FQSGQYIALPRTSRQYQL.....IKYFTSNPPEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:**
1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mhc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp_vertebrate:**
14: sp_unclassified:**
15: sp_virus:**
16: sp_bacteriophage:**
17: sp_archaea:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1784	99.2	360	4	Q9P2Y6
2	1699.5	94.5	423	4	Q15114
3	1694.5	94.2	423	4	Q961Y4
4	1487.5	82.7	422	11	Q9UHR6
5	1477.5	81.6	422	11	Q9QZFO
6	1466.5	81.6	422	11	Q9EOV9
7	916	50.9	198	4	Q9WT18
8	697.5	38.8	416	13	Q9PUF2
9	646.5	36.0	416	6	Q9XSP3
10	637.5	35.5	417	4	Q96B08
11	627	34.9	412	11	P97597
12	627	34.9	417	4	Q96E94
13	537	29.9	279	11	Q9CVD1
14	521.5	29.0	419	6	Q9TVH5
15	514.5	28.6	417	4	Q96ON3
16	514.5	28.6	417	4	Q96A12

17	493.5	27.4	419	4	Q9BS67	Q9BS67 homo sapien
18	477.5	26.6	247	4	Q96KZ9	Q96KZ9 homo sapien
19	463	25.8	1192	5	Q9W475	Q9W475 drosophila
20	461.5	25.7	422	5	Q9VLZ2	Q9VLZ2 drosophila
21	443	24.6	584	5	Q9TZH1	Q9TZH1 caenorhabd
22	440	24.5	424	5	Q9VL87	Q9VL87 drosophila
23	426	23.7	440	5	Q9W478	Q9W478 drosophila
24	409	22.7	424	5	Q9VS66	Q9VS66 drosophila
25	408.5	22.6	419	5	Q9VLZ1	Q9VLZ1 drosophila
26	406.5	22.6	427	5	Q9U9K2	Q9U9K2 aedes aegy
27	406.5	22.6	430	5	Q9VL86	Q9VL86 drosophila
28	403	22.4	323	5	Q19121	Q19121 caenorhabd
29	401	22.3	540	5	Q9TZH2	Q9TZH2 caenorhabd
30	398	22.1	453	5	Q9V342	Q9V342 drosophila
31	392	21.8	423	5	Q61532	Q61532 drosophila
32	391.5	21.8	581	5	Q23318	Q23318 caenorhabd
33	375	20.9	467	5	Q9VCM8	Q9VCM8 drosophila
34	372.5	20.7	455	5	Q9TZC6	Q9TZC6 caenorhabd
35	370	20.6	1430	5	Q9VX86	Q9VX86 drosophila
36	364	20.2	666	5	Q9XU75	Q9XU75 caenorhabd
37	361	20.1	415	5	Q9VS64	Q9VS64 drosophila
38	359	20.0	545	5	Q9V3S6	Q9V3S6 caenorhabd
39	357.5	19.9	400	5	Q9VS67	Q9VS67 drosophila
40	350.5	19.5	354	5	Q9VS68	Q9VS68 drosophila
41	350.5	19.5	418	5	Q961J8	Q961J8 drosophila
42	343	19.1	312	5	Q9VNR3	Q9VNR3 drosophila
43	339	18.9	315	4	Q9NRI9	Q9NRI9 homo sapien
44	331.5	18.4	161	13	Q9OWI2	Q9OWI2 brachydanio
45	328	18.2	418	3	O14418	O14418 metarhizium

ALIGNMENTS

RESULT 1
Q9P2Y6 PRELIMINARY: PRT: 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE CARBOXYPEPTIDASE B-LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL: AB011969; BAA90475.1; -;
DR HSSP: P00730; ZCTC.
DR MEROPS: M14.009; -;
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF002244; Propep_M14; 1.
DR Pfam: PF002246; Zn_carboxypept; 2.
DR PRINTS: PRO0765; CRBOXYPTASEA.
DR SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

Query Match 99.2% Score 1784; DB 4; Length 360;
Best Local Similarity 99.1% Pred. No. 4.8e-155;
Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 FOSGOVLAALPRTSRQOVQONLTITTEYELVMOPTADLIKKQVHFFVNASDVNDVKA 60
DB 23 FOSGOVLAALPRTSRQOVQONLTITTEYELVMOPTADLIKKQVHFFVNASDVNDVKA 82
QY 61 HLNVSGIPCSVLLADVEDLIQOOISNDVSPRASASYEQYHSLNFIYSWIEFITERHPD 120
DB 83 HLNVSGIPCSVLLADVEDLIQOOISNDVSPRASASYEQYHSLNFIYSWIEFITERHPD 142
QY 121 MLTKIHGSSFEKYPPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGH 180
DB 143 MLTKIHGSSFEKYPPLVYLKVSKEQOAKNAIWDGIGHAREMISPAFLMFIIGH 202
QY 181 KRSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPSSEPVKAVASFLRRN 240
DB 203 KRSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPSSEPVKAVASFLRRN 262
QY 241 INQIKAYISMHSYSOHIVFPYSYTRSKSKDHRELVLVASEAVRAIEKTSKNTRYHGHGS 300
DB 263 INQIKAYISMHSYSOHIVFPYSYTRSKSKDHRELVLVASEAVRAIEKTSKNTRYHGHGS 322
QY 301 ETLYLAPGGGDMWYDGLGIKYSFTSNPVEKILPLSLK 338
DB 323 ETLYLAPGGGDMWYDGLGIKYSFTSNPVEKILPLSLK 360
RESULT 2
Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PCPB PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
J. Biol. Chem. 266:21833-21838(1991).
DR EMBL: M75106; AAA60042.1; -.
DR HSP; P00730; IPT.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; PROPEP_M14.
DR InterPro; IPR000834; Zn_Carboxypept.
DR Pfam; PF02244; PROPEP_M14; 1.
DR Pfam; PF00246; Zn_Carboxypept. 1.
DR PRINTS; PR00765; CBBOXPTASEA.
FT CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6E79F4 CRC64;

```

```

Query Match 94.5%; Score 1699.5; DB 4; Length 423;
Best Local Similarity 89.8%; Pred. No. 3, 3e-147;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```

```

QY 1 FOSGOVLAALPRTSRQOVQONLTITTEYELVMOPTADLIKKQVHFFVNASDVNDVKA 60
DB 23 FOSGOVLAALPRTSRQOVQONLTITTEYELVMOPTADLIKKQVHFFVNASDVNDVKA 82
QY 61 HLNVSGIPCSVLLADVEDLIQOOISNDVSPRASASYEQYHSLNFIYSWIEFITERHPD 120
DB 83 HLNVSGIPCSVLLADVEDLIQOOISNDVSPRASASYEQYHSLNFIYSWIEFITERHPD 142
QY 121 MLTKIHGSSFEKYPPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGH 175
DB 323 MLTKIHGSSFEKYPPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGH 175

```

```

DB 143 MLTKIHGSSFEKYPPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGHTOFY 202
QY 176 -----NRMKRNRSPYANNHCIGTDLNRNFASK 203
DB 203 GIIQYTNLRLVDFYVMPVNVGDYDYSMKRNRMRNRSPYANNHCIGTDLNRNFASK 262
QY 204 HMCCEGASSSCSEYTCGLYPSSEPVKAVASFLRRNINQIKAYISMHSYSOHIVFPYSY 263
DB 263 HMCCEGASSSCSEYTCGLYPSSEPVKAVASFLRRNINQIKAYISMHSYSOHIVFPYSY 322
QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYHGHSEETLYLABGGGDMWYDGLIKYSF 323
DB 323 TRSKSDHEELSLVASEAVRAIEKTSKNTRYHGHSEETLYLABGGGDMWYDGLIKYSF 382
QY 324 T 324
DB 383 T 383
RESULT 3
Q061Y4 PRELIMINARY; PRT; 423 AA.
ID Q061Y4
AC Q061Y4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B2 (PLASMA).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKETAL MUSCLE;
RA Strausberg R.;
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC007057; AAH07057.1; -.
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6E7CF5 CRC64;

```

```

Query Match 94.2%; Score 1694.5; DB 4; Length 423;
Best Local Similarity 89.5%; Pred. No. 9, 4e-147;
Matches 323; Conservative 0; Mismatches 1; Indels 37; Gaps 1;

```

```

QY 1 FOSGOVLAALPRTSRQOVQONLTITTEYELVMOPTADLIKKQVHFFVNASDVNDVKA 60
DB 23 FOSGOVLAALPRTSRQOVQONLTITTEYELVMOPTADLIKKQVHFFVNASDVNDVKA 82
QY 61 HLNVSGIPCSVLLADVEDLIQOOISNDVSPRASASYEQYHSLNFIYSWIEFITERHPD 120
DB 83 HLNVSGIPCSVLLADVEDLIQOOISNDVSPRASASYEQYHSLNFIYSWIEFITERHPD 142
QY 121 MLTKIHGSSFEKYPPLVYLKVSKEQOTAKNAIWDGIGHAREMISPAFLMFIIGH 175
DB 143 MLTKIHGSSFEKYPPLVYLKVSKEQOAKNAIWDGIGHAREMISPAFLMFIIGHTOFY 202
QY 176 -----NRMKRNRSPYANNHCIGTDLNRNFASK 203
DB 203 GIIQYTNLRLVDFYVMPVNVGDYDYSMKRNRMRNRSPYANNHCIGTDLNRNFASK 262
QY 204 HMCCEGASSSCSEYTCGLYPSSEPVKAVASFLRRNINQIKAYISMHSYSOHIVFPYSY 263
DB 263 HMCCEGASSSCSEYTCGLYPSSEPVKAVASFLRRNINQIKAYISMHSYSOHIVFPYSY 322
QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYHGHSEETLYLABGGGDMWYDGLIKYSF 323
DB 323 TRSKSDHEELSLVASEAVRAIEKTSKNTRYHGHSEETLYLABGGGDMWYDGLIKYSF 382
QY 324 T 324
DB 383 T 383

```

RESULT 4
ID Q9JH6 PRELIMINARY; PRT; 422 AA.
Q9JH6
AC Q9JH6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE R (THROMBIN-ACTIVATABLE FIBRINOLYSIS INHIBITOR)
DE (1110032P04RIK PROTEIN).
GN CPB2 OR TAFI OR 1110032P04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20341711; PubMed=10878383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not."
RL J. Immunol. 165:1053-1058(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20201966; PubMed=10739389;
RA Marx P.F., Magenar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor."
RL Thromb. Haemost. 83:297-303(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nkaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinchik S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Williams L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsukl S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL, AB021968; BAB03402.1; -;
DR EMBL, AF164524; AAF62385.1; -;
DR EMBL, AK004045; BAB3141.1; -;
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -;
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Proper_M14.
DR InterPro; IPR00834; Zn_carboxept.
DR Pfam; PF02244; Proper_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

Query Match 82.7%; Score 1487.5; DB 11; Length 422;
Best Local Similarity 78.1%; Pred. No. 8,4e-128;
Matches 282; Conservative 20; Mismatches 22; Indels 37; Gaps 1;
QY 1 FOSGOVLAALPRTSRQVQLQNIITTYEIVLMQPVTAADIVKKQVHFVNASDVNVA 60

Db 22 FOSGOVLAALPRTSRQVQLQNIITTYEIVLMQPVTADEFIEKKKEVHFVNASDVSVNA 81
QY 61 HLNVSGIPCSVLADVEDLIQOQISNDTVSPRASASYEQYHSLNFTYSWIEFTERRHPD 120
Db 82 HLNVSRIPFNVLNNVEDLIEQTFNDTVSPRASASYEQYHSLNFTYSWIEFTERRHPD 141
QY 121 MLTKIHIGSSFEKPYLYLVKSGKEQTAKNALWIDCGIHAREWISPAFCIMFYTGQHP 175
Db 142 MLQKIYIGSSFEKPYLYLVKSGKEGRINKNALWIDCGIHAREWISPAFCIMFYTGQHP 201
QY 176 -----NRMRKNSFYANNHCIGTDLNRFPASK 203
Db 202 GKENLYTRLRHVDYFIMPVNVDGYDTWKRMKRMKNSAHKRNRCVGTDLNRFPASK 261
QY 204 HMCSEGASSSCSEYTCGLYSESEPEVKAVASFLRRNINQIKATISMHSOHLTPPYAS 263
Db 262 HMCSEGASSSCSEYTCGLYSESEPEVKAVADFLRRNIDIKATISMHSOHLTPPYAS 321
QY 264 TRSKSDHEELSLVASEAVRAIEKTKNRYTHGSEETLYLAPGGDWIYDLGIKYSP 323
Db 322 NRKSKDHLESLVASEAVRAIESINKNRYTHGSGESLYLAPGSDWIDYDLGIKYSP 381
QY 324 T 324
Db 382 T 382
RESULT 5
Q9QZF0 PRELIMINARY; PRT; 422 AA.
Q9QZF0
AC Q9QZF0;
ID Q9QZF0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE U.
GN CPB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAFJ;
He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
RT gene."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF186188; AAF00528.1; -;
DR HSSP; P00730; 5CPA.
DR MEROPS; M14.009; -;
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Proper_M14.
DR InterPro; IPR00834; Zn_carboxept.
DR Pfam; PF02244; Proper_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
SQ SEQUENCE 422 AA; 48832 MW; B53FFFB09943954E CRC64;

Query Match 82.2%; Score 1477.5; DB 11; Length 422;
Best Local Similarity 77.6%; Pred. No. 6,9e-127;
Matches 280; Conservative 21; Mismatches 23; Indels 37; Gaps 1;
QY 1 FOSGOVLAALPRTSRQVQLQNIITTYEIVLMQPVTAADIVKKQVHFVNASDVNVA 60
Db 22 FOSGOVLAALPRTSRQVQLQNIITTYEIVLMQPVTADEFIEKKKEVHFVNASDVSVNA 81
QY 61 HLNVSGIPCSVLADVEDLIQOQISNDTVSPRASASYEQYHSLNFTYSWIEFTERRHPD 120
Db 82 HLNVSRIPFNVLNNVEDLIEQTFNDTVSPRASASYEQYHSLNFTYSWIEFTERRHPD 141
QY 121 MLTKIHIGSSFEKPYLYLVKSGKEQTAKNALWIDCGIHAREWISPAFCIMFYTGHP 175

```

Db 142 MLOKTIIGSSFEKYPYLKVSKEQRIKNAIMIDGCHAREWISPAFLMFGYTOFH 201
QY 176 -----NRMRKNSFYANNHCIGTDLNRFPASK 203
Db 202 GKENTYRLRLRHVDYIMPMNVNDGYDTYMKRMRKRNRSAHKNRRCGTDLNRFPASK 261
QY 204 HMCCEGASSSCSETCYGLYPESEPEVKAVASFLRRNIQIKAYISMHSYSHIYFPYSY 263
Db 262 HMCCEGASSSCSETCYGLYPESEPEVKAVADFLRRNIHIAKAYISMHSYSHIYFPYSY 321
QY 264 TRSKSDHELSLVAEAVRAIEKTSKNTRYTHGSESLYLAPGGDDWITDGIKYSF 323
Db 322 NRSKSDHELSLVAEAVRAIEKTSKNTRYTHGSESLYLAPGGDDWITDGIKYSF 381
QY 324 T 324
Db 382 T 382

RESULT 6
Q9EUV9 PRELIMINARY: PRT: 422 AA.
AC Q9EUV9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PRE-PROCARBOXYPEPTIDASE R.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20471387; PubMed=11021404;
RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA Okada N., Okada H.;
RT "Molecular cloning and partial characterization of rat
RT procarboxypeptidase R and carboxypeptidase N.";
RL Microbiol. Immunol. 44:719-728(2000).
DR EMBL; AB042598; BAB18617.1; -.
DR HSSP; P00730; ZCTC.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
SQ SEQUENCE 422 AA; 48826 MW; FFFD32A51A9366C8 CRC64;

Query Match 81.6%; Score 1466.5; DB 11; Length 422;
Best Local Similarity 76.7%; Pred. No. 7e-126;
Matches 277; Conservative 22; Mismatches 25; Indels 37; Gaps 1;

QY 1 FOSGOVLAALPRTSRQOVQVONLTITTEIYVLMQPVTAADLVKKQVHFFYNASDVNVKA 60
Db 22 FOSGHVLSALPRTSRQOVQVONLTITTEIYVLMQPVTAADLVKKQVHFFYNASDVNVKA 81
QY 61 HLNVSIGPCSVLADVEDLIQOQISNDTVSPRASASYEQYHSLNIEYSWIEFTERHPD 120
Db 82 YLNASRIPEFVNLNNVBDLIQOQISNDTVSPRASASYEQYHSLNIEYSWIEFTERHPD 141
QY 121 MLTKIHGSSFEKYPYLKVSKEQTAKNNAIWDGCHAREWISPAFLMFGYTOFH 175
Db 142 MLOKTIIGSSFEKYPYLKVSKEQTAKNNAIWDGCHAREWISPAFLMFGYTOFH 201
QY 176 -----NRMRKNSFYANNHCIGTDLNRFPASK 203
Db 202 GKENTYRLRLRHVDYIMPMNVNDGYDTYMKRMRKRNRSAHKNRRCGTDLNRFPASK 261
QY 204 HMCCEGASSSCSETCYGLYPESEPEVKAVASFLRRNIQIKAYISMHSYSHIYFPYSY 263

```

```

Db 262 HMCCEGASSSCSETCYGLYPESEPEVKAVADFLRRNIHIAKAYISMHSYSHIYFPYSY 321
QY 264 TRSKSDHELSLVAEAVRAIEKTSKNTRYTHGSESLYLAPGGDDWITDGIKYSF 323
Db 322 NRSKSDHELSLVAEAVRAIEKTSKNTRYTHGSESLYLAPGGDDWITDGIKYSF 381
QY 324 T 324
Db 382 T 382

RESULT 7
Q9NT18 PRELIMINARY: PRT: 198 AA.
AC Q9NT18;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BA139H14.2 (CARBOXYPEPTIDASE B2 (PLASMA)) (FRAGMENT).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TROMANS A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL37141; CAB92622.1; -.
DR HSSP; P00730; ICPX.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
FT NON_TER 198
SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;

Query Match 50.9%; Score 916; DB 4; Length 198;
Best Local Similarity 99.4%; Pred. No. 6.2e-76;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FOSGOVLAALPRTSRQOVQVONLTITTEIYVLMQPVTAADLVKKQVHFFYNASDVNVKA 60
Db 23 FOSGOVLAALPRTSRQOVQVONLTITTEIYVLMQPVTAADLVKKQVHFFYNASDVNVKA 82
QY 61 HLNVSIGPCSVLADVEDLIQOQISNDTVSPRASASYEQYHSLNIEYSWIEFTERHPD 120
Db 83 HLNVSIGPCSVLADVEDLIQOQISNDTVSPRASASYEQYHSLNIEYSWIEFTERHPD 142
QY 121 MLTKIHGSSFEKYPYLKVSKEQTAKNNAIWDGCHAREWISPAFLMFGYTOFH 175
Db 143 MLOKTIIGSSFEKYPYLKVSKEQTAKNNAIWDGCHAREWISPAFLMFGYTOFH 201

RESULT 8
Q9PUF2 PRELIMINARY: PRT: 416 AA.
AC Q9PUF2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE HOMOLOQ.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_Taxid=8724;
RN [1]
RP SEQUENCE FROM N.A.
RA MURDACH A.F., HAYASHI M.A.F., CAMARGO A.C.M.;

```



```

Db 22 FESEKVFRRVNEDEENINIRLASTQTQIDFMKPSVTOIKPHSTVDKRVKADVTVEN 81
QY 61 HLNVSIGIPCSVLLADVEDLIDQOISNDTVSPRASASYEQYHSNLEYISWIEFITERHPD 120
Db 82 VIKONELQKVKILISNLRNVEAOFDSRV---RATGSHYERKNMKEITEAMTQOVATENRA 138
QY 121 MLTKIHGSSFEKYPYLYLVKSGKEQTAKNAIWDGIGHAREWISPAFLCMTF----- 173
Db 139 LISRSVIGTFEGRAIYLLKV-GKAGQNKPAJFMDCGFHAREWISPAFCQMFVREAVRTY 197
QY 174 -----GH-----NRMKRNRSPYANNHCIGTDLNRNFASK 203
Db 198 GREIQVTELLDKIDFVLPVNLIDGIYTWTSKRFNRKTRSTHTGSSCIGTDPNRNF-DA 256
QY 204 HMCBEGASSSCSEYTCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYQHIYFPPSY 263
Db 257 GICEIGASHPNCPDEYTCGPAESEKETKALADPFRNKLISSIKAYILHSYQMLIYPPSY 316
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWYIDLGIRYSF 323
Db 317 AKKLENNALNAKATVKEL-ASLHGTYKYTGPGATTIYPAAAGSDOMAYDQIRYSF 375
QY 324 T 324
Db 376 T 376

```

```

RESULT 11
ID P97597 PRELIMINARY; PRT: 412 AA.
AC P97597;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MAST CELL CARBOXYPEPTIDASE A PRECURSOR (FRAGMENT).
GN R-CPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY;
RX MEDLINE=97149430; PubMed=8996238;
RA Lutzelschab C., Pejler G., Aveskogh M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations."
RL J. Exp. Med. 185:13-29(1997).
DR EMBL: U67914; AAB48267.1; -.
DR HSSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR InterPro: IPR001146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypep.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypep; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KM signal; Carboxypeptidase.
FT NON_TER 1
FT STGNL 1
FT CHAIN 11 412 POTENTIAL.
SQ SHOUNCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

```

```

Query Match 34.9%; Score 627; DB 11; Length 412;
Best Local Similarity 37.5%; Pred. No. 5e-49;
Matches 135; Conservative 55; Mismatches 128; Indels 42; Gaps 6;
QY 1 FOSGOVLAALPRTSRQOVQVONLTITTEIYVLMQPVATDLVKKQVHFVNADVDNVKA 60
Db 16 FDEKVFRRVRLQDEKQASLUKNLTQTEIDFWPDAIHDAIVNMTVDKRVTEREKESQIOS 75

```

```

QY 61 HLNVSIGIPCSVLLADVEDLIDQOISNDTVSPRASASYEQYHSNLEYISWIEFITERHPD 120
Db 76 TLEQHKMDYELIINLQOEIDKQF--DYKEELAGHSHSAKYNNDKNIYSWTEKWEKHE 133
QY 121 MLTKIHGSSFEKYPYLYLVKSGKEQTAKNAIWDGIGHAREWISPAFLCMTF----- 173
Db 134 MYSRIKISTYEDNPNLYLKI-GRDGERKALFMDCGIGHAREWISPAFCQMFVQAAKSY 192
QY 174 GHN-----GH-----NRMKRNRSPYANNHCIGTDLNRNFASK 203
Db 193 GKNKIMTKLDRMNFYLPVENVDSYIWSWTKRMRNRKRNKNSNCIGTDLNRNF-DV 251
QY 204 HMCBEGASSSCSEYTCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYQHIYFPPSY 263
Db 252 SWDSSPNTDNCISLYRGPAPSEKETKAVNFTSRHLSNSTIKAYITTHYSQMLIYPPSY 311
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWYIDLGIRYSF 323
Db 312 TIKLPNHQDLKVRATDVL-SSRYETRYIYGPASTIYKTSGLSSLDMAVYDGIKHTF 370

```

```

RESULT 12
ID Q96E94 PRELIMINARY; PRT: 417 AA.
AC Q96E94;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SIMILAR TO CARBOXYPEPTIDASE A3 (MAST CELL).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW, AND ACUTE MYELOCYTIC LEUKEMIA;
RA Strimberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012613; AAH12613.1; -.
KM Carboxypeptidase.
SQ SEQUENCE 417 AA; 48714 MW; F781EF23F570E98 CRC64;

```

```

Query Match 34.9%; Score 627; DB 4; Length 417;
Best Local Similarity 35.7%; Pred. No. 5.1e-49;
Matches 129; Conservative 71; Mismatches 117; Indels 44; Gaps 7;
QY 1 FOSGOVLAALPRTSRQOVQVONLTITTEIYVLMQPVATDLVKKQVHFVNADVDNVKA 60
Db 21 FDEKVFRRVRLQDEKQADIIKDLAKTNEIDFWPCTHHVAAKMAYDFRVERESQAIOS 80
QY 61 HLNVSIGIPCSVLLADVEDLIDQOISNDTVSPRASASYEQYHSNLEYISWIEFITERHPD 119
Db 81 ALDQNMKHYYIILHDQEELEKQFDVKEDIPGRHS--YAKINNMKEIYAWTEKMKMDKYP 137
QY 120 MLTKIHGSSFEKYPYLYLVKSGKEQTAKNAIWDGIGHAREWISPAFLCMTF----- 173
Db 138 EMYSRIKISTYEDNPNLYLKI-GEKNERRKALFMDCGIGHAREWISPAFCQMFVQAAKSY 196
QY 174 -----GH-----NRMKRNRSPYANNHCIGTDLNRNFASK 202
Db 197 YGKNKIMTKLDRMNFYLPVENVDSYIWSWTKRMRNRKRNKNSNCIGTDLNRNF-N 255
QY 203 HMCBEGASSSCSEYTCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYQHIYFPPSY 262
Db 256 ASWNSIPNTNDPCADNRYRSAPSEKETKAVNFTSRHLSNSTIKAYITTHYSQMLIYPPYG 315
QY 263 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWYIDLGIRYSF 322
Db 316 YTSKLPNHNEDIAKAKIGTDVL-STRYETRYIYGPASTIYKTSGLSSLDMAVYDGIKHT 374
QY 323 F 323

```


DB 375 F 375

RESULT 13

OC 09CVD1 PRELIMINARY; PRT: 279 AA.
AC 09CVD1
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 221008M23RIK PROTEIN (FRAGMENT).
GN 221008M23RIK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Felschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Rang B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK008678; BAB2826.1; -
DR HSSP: P09955; 1NSA.
DR MGD: MGI:1923953; 2210008M23RIK.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; UNKNOWN_1.
FT NON_TER 279
SQ SEQUENCE 279 AA; 31970 MW; 29468AFA0B67F0AA CRC64;

Query Match 29.9%; Score 537; DB 11; Length 279;
Best Local Similarity 46.0%; Pred. No. 5e-41; Mismatches 51; Indels 40; Gaps 4;
Matches 104; Conservative 31;

QY 136 LVYLVKSGKEGTAKNAIWDGTHAREWISPAFCIMFT----- 173
DB 1 MYVLKI-GDPRNKPFAIFIDCGFHAREWISPAFCQMFVEAVTYRQETIMKRLDELDF 59
QY 174 -----GHNMRKNSRFYANNHCIGTDLNRNPAKSWCEGASSSCSET 218
DB 60 VYLPVNVNIDGYTYWAKDMMWRKTRSTAGYSCFQVDPNKNF-DAGWCEVGASRSCSDT 118
QY 219 YGGLPESPEPEYKAVASFRRNINOKAYISMYSOHIVPEPYSTRSKSDEELSLVA 278
DB 119 YGGLPESPEPEKATADFRONLSIKATUTYHYSQMMILPYSYKLPENEEELNALV 178
QY 279 SEAVRAIEKTSKNTRYTHGHSEETLYLAPGGGDDWITDGIKYSFT 324
DB 179 KGAARKEL-STLHGKTYTYPGATTYIPAGAGSDMWAYDQGIKYSFT 223

RESULT 14

OC 09TV85 PRELIMINARY; PRT: 419 AA.
AC 09TV85;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE A1 (EC 3.4.17.1).
GN PCPAP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9823;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99192816; PubMed=10092856;
RA Darnis S., Juge N., Marinho C., Aviles F.X., Puigserver A., Chaux J.C.,
RA Guo X.J.;
RT "Cloning, sequencing and functional expression of a cDNA encoding
porcine pancreatic preprocarboxypeptidase A1."
RL Eur. J. Biochem. 259:719-725(1999).
DR EMBL: AF076222; AAD17690.1; -
DR HSSP: P09954; 1PCA.
DR MEROPS: M14.001; -
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14. 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase.
SQ SEQUENCE 419 AA; 47235 MW; 84B4CB557B714FC1 CRC64;

Query Match 29.0%; Score 521.5; DB 6; Length 419;
Best Local Similarity 34.2%; Pred. No. 2.3e-39;
Matches 127; Conservative 61; Mismatches 124; Indels 59; Gaps 11;

QY 1 FQSGQVLAALPRTSRQVQVQNTLTTEYIVL--WQ-PVTADLVKKQVHFFVNASPDVN 57
DB 20 FVGHQVLRISVDEDAQYQKKELEHLDQDWRGARGCFPIQV-----VPPSIOA 74
QY 58 VKAHLNVSGIPCSVLADVEDLI---QQOISNDVSPRASASY-YEQYHSLNETYSWIEF 113
DB 75 VKVFLHNGIRYTIMEDVOLLDEDEQOMFASQGRARTSTFNATYATHTLEIYDEMID 134
QY 114 ITRHRPMLKIHIGSFEKYPPLVLYKSGKEOTAKY---AIVTDCIHAREWISPAFC 170
DB 135 LVAEHPQLVSKLDIGSSYEGRPYLKFS---TGGNNRPALWIDGISHREWVTOASGV 190
QY 171 WFI-----GHNMRKNSRFYANNHCIG 193
DB 191 WFAKKTITEDGQDPATFALINDIDPLEIYTNPDGFAFTSENMRKTKTSRISGSCV 250
QY 194 TDLNRNPAKSWCEGASSSCSETYCGLYPESPEYKAVASFLRNINOKAYISMYS 253
DB 251 VDPNRNM-DAGFGAGASNPSCSETYHGKPPNSEVEYKSYVDVNDGN- IKAFTISHY 308
QY 254 SOHIVPEPYSTRSKSDEELSLVAEAVRAIEKTSKNTRYTHGHSEETLYLAPGGDDW 313
DB 309 SOLILVPGYGTAPADKDELDOISKSAVAL-TSLVGTFRFGYSITITTYQASGGTIDW 367
QY 314 IYDGIKYSFT 324
DB 368 TYNQGIKYSFS 378
RESULT 15
OC 09GON3 PRELIMINARY; PRT: 417 AA.
AC 09GON3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:45:13 ; Search time 65 Seconds

(without alignments)
420.372 Million cell updates/sec

Title: us-09-980-881-4

Perfect score: 1338
Sequence: 1 AASAYEQHSLNEITSWIE.....IKYTSFSPVEKILPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Genesec_032802:*

- 1: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/genesec/genesecp-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/genesec/genesecp-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/genesec/genesecp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1338	100.0	246	21	Human brain carbox
2	1338	100.0	338	21	Human brain carbox
3	1325	99.0	360	21	Human brain carbox
4	1239.5	92.6	423	17	Human plasma carbox
5	1239.5	92.6	423	18	Human plasma carbox
6	1239.5	92.6	423	20	Human plasma carbox
7	1234.5	92.3	423	14	Human plasma carbox
8	816.5	61.0	211	21	Human plasma carbox
9	612.5	45.6	324	22	Human plasma carbox
10	610	45.6	307	17	Human plasma carbox
11	582	43.5	306	16	Human plasma carbox

12	582	43.5	306	22	AAW04477
13	582	43.5	329	16	AAW75131
14	567	42.4	349	17	AAW06174
15	567	42.4	349	17	AAW06173
16	567	42.4	349	17	AAW06172
17	567	42.4	349	17	AAW06171
18	567	42.4	349	17	AAW06170
19	560	41.9	349	17	AAW06169
20	560	41.9	349	17	AAW06168
21	559	41.8	349	17	AAW06167
22	559	41.8	349	17	AAW06166
23	556	41.6	349	18	AAW13750
24	554	41.4	349	18	AAW13752
25	554	41.4	349	18	AAW13751
26	553	41.3	349	18	AAW13750
27	552	41.3	349	18	AAW13753
28	551	41.2	349	18	AAW13754
29	551	41.2	349	18	AAW13755
30	551	41.2	349	18	AAW13756
31	551	41.2	349	18	AAW13757
32	550.5	41.1	354	22	AAW13758
33	550.5	41.1	354	22	AAW13759
34	550.5	41.1	362	22	AAW13760
35	550.5	41.1	374	22	AAW13761
36	550.5	41.1	374	22	AAW13762
37	549	41.0	349	18	AAW13763
38	549	41.0	349	18	AAW13764
39	549	41.0	349	18	AAW13765
40	549	41.0	349	18	AAW13766
41	549	41.0	349	18	AAW13767
42	549	41.0	349	18	AAW13768
43	548	41.0	349	18	AAW13769
44	546	40.8	349	18	AAW13770
45	492	36.8	399	22	AAW13771

ALIGNMENTS

RESULT 1	AAW1459	standard; Protein; 246 AA.
XX	AAW1459;	
XX	AAW1459;	
DT	01-MAR-2001 (first entry)	
XX	Human brain carboxypeptidase B protein SEQ ID NO 4.	
DE	Human brain carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase; treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine; cerebroprotective; antialzheimer's; neurotrophic; neuroprotective; hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage; Down's syndrome; head trauma.	
KW	Homo sapiens.	
OS	Homo sapiens.	
XX	WO200066717-A1.	
XX	09-NOV-2000.	
PD	01-MAY-2000; 2000MO-JP02878.	
XX	30-APR-1999; 99JP-0125169.	
PR	(MATS/) MATSUMOTO A.	
PA	Matsumoto A;	
XX	WPI; 2000-687534/67.	
XX	Human brain carboxypeptidase B isolated from the hippocampus useful for screening agents for the treatment of Alzheimer's and other brain	
PT		

PT disorders -
 PS Disclosure; Page 74-75; 84pp; Japanese.
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity, and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 SQ Sequence 246 AA;

Query Match 100.0%; Score 1338; DB 21; Length 246;
 Best Local Similarity 100.0%; Pred. No. 5.3e-134;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASASYEQYHSLNELYSWIEFTEHRHPDMLTKIHGSSFEKPYLVKVGKEQTAKNAI 60
 DB 1 asasyeqyhslnelyswiefterhpdmltkihgssfekpylvkvsgkeqtaknai 60
 QY 61 WIDGIIHREWISPAFCIMFIGNHMRKNSFYANNHCIGTDLNRNFASKHMCCEGASS 120
 DB 61 widgiiharewispafciwfighnmwrknsfyannhcigtldlnrnfaskhmccegass 120
 QY 121 SSCSEYCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYSOHIVFPYSYTRSKSDHE 180
 DB 121 sscseycglypesepevkavasflrrninqikayismhsysohivfypsyrtskskde 180
 QY 181 ELISIVASEAVRAIEKTSKNTRYTHGHSEETLYLAPGGDDMIYDLGKYSFTSNPVEKL 240
 DB 181 elisivaseavraiektskntrythghsetlylapggddwyldgkysftsnpvekl 240
 QY 241 LPLSLK 246
 DB 241 lplslk 246

RESULT 2
 AAB11458
 ID AAB11458 standard; Protein; 338 AA.
 AC AAB11458;
 XX
 DT 01-MAR-2001 (first entry)
 XX
 DE Human brain carboxypeptidase B protein SEQ ID NO 3.
 XX
 DE Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KM treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KM cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KM hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KM Down's syndrome; head trauma.
 XX
 OS Homo sapiens.
 XX
 PN WO2000066717-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 01-MAY-2000; 2000WO-JP02878.
 XX
 PR 30-APR-1999; 99JP-0125169.
 XX
 PA (MATS/) MATSUMOTO A.
 XX
 PI Matsumoto A;
 XX
 DR WPI; 2000-687534/67.

XX Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 PS Disclosure; Page 71-73; 84pp; Japanese.
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity, and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 SQ Sequence 338 AA;

Query Match 100.0%; Score 1338; DB 21; Length 338;
 Best Local Similarity 100.0%; Pred. No. 8.6e-134;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASASYEQYHSLNELYSWIEFTEHRHPDMLTKIHGSSFEKPYLVKVGKEQTAKNAI 60
 DB 93 asasyeqyhslnelyswiefterhpdmltkihgssfekpylvkvsgkeqtaknai 152
 QY 61 WIDGIIHREWISPAFCIMFIGNHMRKNSFYANNHCIGTDLNRNFASKHMCCEGASS 120
 DB 61 widgiiharewispafciwfighnmwrknsfyannhcigtldlnrnfaskhmccegass 120
 QY 153 WIDGIIHREWISPAFCIMFIGNHMRKNSFYANNHCIGTDLNRNFASKHMCCEGASS 212
 DB 153 widgiiharewispafciwfighnmwrknsfyannhcigtldlnrnfaskhmccegass 212
 QY 121 SSCSEYCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYSOHIVFPYSYTRSKSDHE 180
 DB 213 sscseycglypesepevkavasflrrninqikayismhsysohivfypsyrtskskde 272
 QY 181 ELISIVASEAVRAIEKTSKNTRYTHGHSEETLYLAPGGDDMIYDLGKYSFTSNPVEKL 240
 DB 273 elisivaseavraiektskntrythghsetlylapggddwyldgkysftsnpvekl 332
 QY 241 LPLSLK 246
 DB 333 lplslk 338

RESULT 3
 AAB11457
 ID AAB11457 standard; Protein; 360 AA.
 AC AAB11457;
 XX
 DT 01-MAR-2001 (first entry)
 XX
 DE Human brain carboxypeptidase B protein.
 XX
 DE Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KM treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KM cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KM hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KM Down's syndrome; head trauma.
 XX
 OS Homo sapiens.
 XX
 PN WO2000066717-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 01-MAY-2000; 2000WO-JP02878.
 XX
 PR 30-APR-1999; 99JP-0125169.
 XX
 PA (MATS/) MATSUMOTO A.
 XX
 PI Matsumoto A;
 XX
 DR WPI; 2000-687534/67.

PI Matsumoto A;
XX
XX WPI, 2000-687534/67.
DR N-PSDB; AAC81962.
DR
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
XX
PS Claim 1; Page 68-71; 84pp; Japanese.
XX
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC neurotropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
XX Sequence 360 AA;

	Query Match	99.0%: Score 1325; DB 21; Length 360;
	Best Local Similarity	98.8%: Pred. No.2.3e-132;
	Matches 243; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
QY	1 ASASYEEQYHSLNLTYSNIETFERRHDPMLTKRIHGSEFEKPYLVYLKVGSKGEOTAKNAI 60	
DB	115 aaasyeyegyhslnelysylefiterhpmnlkthhgsgsfekyplvylkvsygeqtaknai 174	
QY	61 WIDGIIHAREWISPAFCIMFIGNBMRKNSRYANNHCIQTDILNRNPAASKHWCEGASS 120	
DB	175 widcgiharewispafciwfighnmwrknrsfyannhoicgtdlnsnfyskhwceagass 234	
QY	121 SSCSTYTCGLVYESPEPKAVASFRRININQKATISMHSYQHIYFPYSTRSKSDHE 180	
DB	235 sscsetyqglypespevkavasflrrlnnqlkayismhsyqhiyfypyslrskskdhe 294	
QY	181 ELSLVASEAVRAIREKTSKNTRYTHGSGSETLYLAGGGDDVYDGIKSPSNPVEKL 240	
DB	295 elslvasaeavraidktskntlytnhgseelylapggddwlydglkysfscsnppvekl 354	
QY	241 LPLSLK 246	
DB	355 lplslk 360	
RESULT	4	
AA90293	AA90293	
ID	AA90293 standard; Protein; 423 AA.	
XX	AA90293;	
AC	AA90293;	
XX	12-APR-1996 (first entry)	
DT	12-APR-1996 (first entry)	
XX	Human plasma carboxypeptidase B.	
DE	Human plasma carboxypeptidase B.	
XX	Plasma carboxypeptidase B; hpcPB; antibody; detection;	
KM	purification; plasminogen; affinity column.	
XX		
OS	Homo sapiens.	
XX		
FM	Key	Location/Qualifiers
FT	Active-site	181
FT	/note=	"Residue suspected of being involved in catalytic activity of enzyme."
FT	Binding-site	183
FT	/note=	"Residue suspected of being part of the binding site of enzyme."
FT	Binding-site	256..257
FT	/note=	"Residues suspected of being part of the binding site of enzyme."

FT	Active-site	310	binding site of enzyme."
FT		/note=	"Residue suspecting of being involved in
FT			catalytic activity of enzyme."
FT	Binding-site	312	
FT		/note=	"Residue suspected of being part of the
FT			binding site of enzyme."
FT	Active-site	369	
FT		/note=	"Residue suspecting of being involved in
FT			catalytic activity of enzyme."
FT	Binding-site	370..371	
FT		/note=	"Residues suspected of being part of the
FT			binding site of enzyme."
FT	Binding-site	394	
FT		/note=	"Residue suspected of being part of the
FT			binding site of enzyme."

PN US5474901-A.
XX
PD 12-DEC-1995.
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
XX
PR 14-OCT-1992; 92US-0959944.
XX
PR 15-DEC-1993; 93US-0167727.
XX
PR 19-JUL-1994; 94US-0277540.
XX
PA (GETH) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
DR WPI; 1996-039508/04.
XX
DR N-PSDB; AAT11671.
XX
XX Antibody to human plasma carboxypeptidase B - useful for detecting
PI and purifying hPCPB for use in treating clotting disorders e.g.
PT haemophilia A
XX
XX Disclosure; Figure 4; 40pp; English.
PS
XX
XX An antibody which specifically binds human plasma carboxypeptidase B
CC (hPCPB) and does not cross react with other carboxypeptidases is
CC useful for the detection of hPCPB in vitro. The antibody is also
CC used for purifying hPCPB from a sample. Purification comprises
CC passing a sample thought to contain hPCPB over either a column to
CC which antibody has been bound, or a plasminogen affinity column,
CC eluting the column and then recovering the fraction containing the
CC hPCPB.
XX
SQ Sequence 423 AA;

```

OY 204 HGHGSETLYLAPGGDDWYDGIKYSFT 232
PT |||||||||||||||||||||||||||||||
DB 355 hghgsetcylapggddwlydlygkysft 383

RESULT 5
AAW14733
ID AAW14733 standard; Protein: 423 AA.
XX
AC AAW14733;
XX
DT 08-MAY-1997 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
XX plasma; plasminogen.
XX
OS Homo sapiens.
XX
FH Key
FH Peptide
FT 1..22 Location/Qualifiers
FT /note= "Signal peptide"
FT 23..423
FT /note= "Mature PCPB"
FT 114
FT /note= "Potential clip site for activation of PCPB
FT as a carboxypeptidase"
FT 181
FT /note= "Expected to be involved in catalytic activity"
FT 183
FT /note= "Expected to be involved in substrate binding"
FT 184
FT /note= "Expected to be involved in catalytic activity"
FT 256
FT /note= "Expected to be involved in substrate binding"
FT 257
FT /note= "Expected to be involved in substrate binding"
FT 310
FT /note= "Expected to be involved in catalytic activity"
FT 312
FT /note= "Expected to be involved in catalytic activity"
FT 369
FT /note= "Expected to be involved in substrate binding"
FT 370
FT /note= "Expected to be involved in catalytic activity"
FT /note= "Expected to be involved in substrate
FT binding, thought to determine specificity
FT of PCPB as a carboxypeptidase B"
FT 371
FT /note= "Expected to be involved in substrate binding"
FT 394
FT /note= "Expected to be involved in substrate binding"
XX
PN USS593674-A.
XX
PD 14-JAN-1997.
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
PR 27-APR-1995; 95US-0430787.
XX
PA (GETH ) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
DR WPI; 1997-099413/09.
DR N-PSDB; AAT62846.
XX
PT Using human plasma carboxy:peptidase B in blood coagulation - 1s

```

```

PT functionally related to carboxy:peptidase A and pancreas
PT carboxy:peptidase B
XX
XX Example 2: Column 37-42; 39pp; English.
XX
XX This sequence represents human plasma carboxypeptidase B (PCPB) which
XX has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
XX PCPB may be used therapeutically in haemostatic regulation. PCPB is
XX purified from human plasma or by transformed cell culture by
XX extraction using plasminogen bound to a solid phase.
XX
SO Sequence 423 AA;

```

```

Query Match 92.6%; Score 1239.5; DB 18; Length 423;
Best Local Similarity 86.2%; Pred. No. 3,7e-123;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```

```

OY 1 ASASYEQYHSLNDEIYSEIFTERHHPDMLTKIHGSSFEKYPPLVAKSGKEOTAKNAI 60
DB |||||||||||||||||||||||||||||||
DB 115 asaayegyhlnelylwlefterhpdmltklhgsfeykypplvlyksgkeqtaknal 174
OY 61 WIDCGIAREWISPAFLCMTFIGH-----
DB |||||||||||||||||||||||||||||||
DB 175 widcgiharewispaflcwtfghlqfygylgqytnllrlvdfympvvnvdgydyswk 234
OY 84 NRMWRKRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSEYCGIYPSPEEVKAVAS 143
DB |||||||||||||||||||||||||||||||
DB 235 nrmwrkrsfyannhcgtdlnrnfnaskhwccegaassscseycgilypspeevkavas 294
OY 144 FLRRNIQIKAYISMHSYSGHIVFPYSYTRSKSDHELSLVASEAVRALEKTSKNTRYT 203
DB |||||||||||||||||||||||||||||||
DB 295 flrrniqikayismhsyqghivfpyysytrskskdheelslvaseavratelktskntlyt 354
OY 204 HGHGSETLYLAPGGDDWYDGIKYSFT 232
DB |||||||||||||||||||||||||||||||
DB 355 hghgsetcylapggddwlydlygkysft 383

```

```

RESULT 6
AAW92270
ID AAW92270 standard; Protein: 423 AA.
XX
AC AAW92270;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW polymorphism detection; thrombotic disease.
XX
OS Homo sapiens.
XX
FH Key
FH Peptide
FT 1..22 Location/Qualifiers
FT /note= "signal peptide"
FT 23..423
FT /note= "mature PCPBthr147"
XX
PN W09855645-A1.
XX
PD 10-DEC-1998.
XX
PF 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD ) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
DR WPI; 1999-045800/04.

```

DR N-PSDB: AAV74302.
 XX
 PT Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 PT humans
 XX
 PS Disclosure: Page 25-26; 35pp; English.
 XX
 CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
 CC hPCPBtr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphs of the protein
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.
 CC
 XX
 SQ Sequence 423 AA:
 Query Match 92.6%; Score 1239.5; DB 20; Length 423;
 Best Local Similarity 86.2%; Pred. No. 3.7e-123;
 Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
 QY 1 ASASYEYQYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPYLYLKYSKGKQTAKNAI 60
 DB 115 asasyeqyhslnelyswiefierhpdmltkihgssfekyplylylkysgkeqtaknal 174
 QY 61 WTIDCGIHAREWISPAFLCWFIGH----- 83
 DB 175 wldcgiharewispafclwfightlqfygllygqyltlldvlympvvnvdgdydyswkk 234
 QY 84 NRMWRKRSFYANNHCIGDNLNRNFASKHMCCEGASSSCSEFYCGLYPSEPEPVAVAS 143
 DB 235 nrmwrkrfsfyanhncigtdlnrnrfaskhmccegassscsefycglypsepepvavavas 294
 QY 144 FLRRNINQKATYSMSYSGOHVFPYSTRSKSKDHEELVASEAVRAIEKTSKTRIT 203
 DB 295 flrrninqkayismysgohvfpysytrskskdheelslvaseavraiektstknrltyt 354
 QY 204 HGHGSETLYAPGGDDWYIDGIRKYSFT 232
 DB 355 hghgsetlylapggddwlyldgirkysft 383
 RESULT 7
 AAR36273
 ID AAR36273 standard; Protein: 423 AA.
 AC AAR36273;
 XX 24-AUG-1993 (first entry)
 XX Human plasma carboxypeptidase B.
 DE PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
 KM Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..22
 FT /label= signal_peptide
 FT Cleavage-site 114
 FT /note= "putative site of cleavage by trypsin to
 FT activate PCPB as a carboxypeptidase"
 FT Active-site 181
 FT /note= "expected to be involved in catalytic

FT Binding-site 183 activity"
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 184
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 256..257
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 310
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 312
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 369
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 370..371
 FT /note= "expected to be involved in substrate
 FT binding; Asp 370 is the residue which
 FT determines specificity of the PCPB as
 FT a carboxypeptidase B"
 FT Binding-site 394
 FT /note= "expected to be involved in substrate
 FT binding"
 FT US5206161-A.
 XX 27-APR-1993.
 PD 01-FEB-1991; 91US-0649591.
 PF 01-FEB-1991; 91US-0649591.
 PR 01-FEB-1991; 91US-0649591.
 PA (GETH) GENENTECH INC.
 XX Drayna DT, Eaton DL;
 PI WPI: 1993-151724/18.
 DR N-PSDB; AA041001.
 XX
 PT New human plasma carboxypeptidase B - used as haemostatic
 PT regulator for clotting blood, partic. for treating blood clotting
 PT disorders, e.g. haemophilia
 PS Claim 2; Fig 4; 40pp; English.
 XX
 CC This amino acid sequence was deduced from a clone isolated from a
 CC human liver cDNA library. There is about 40% sequence identity
 CC between the deduced preproPCPB amino acid sequence and pro-human
 CC mast cell carboxypeptidase A and between preproPCPB and prepro-rat
 CC carboxypeptidase B. Human PCPB has the same substrate binding sites
 CC as, and shares 6 cysteine residues which form 3 intramolecular
 CC disulphide bonds with, bovine and rat carboxypeptidase B. The
 CC presence of the same amino acid (Asp at position 348 of the mature
 CC protein) at the region in carboxypeptidases that determines
 CC substrate specificity, suggests that PCPB represents a plasma-
 CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
 CC tPA of plasminogen to plasmin in the presence of fibrinogen.
 CC
 XX
 SQ Sequence 423 AA:
 Query Match 92.3%; Score 1234.5; DB 14; Length 423;
 Best Local Similarity 85.9%; Pred. No. 1.3e-122;
 Matches 231; Conservative 0; Mismatches 1; Indels 37; Gaps 1;
 QY 1 ASASYEYQYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPYLYLKYSKGKQTAKNAI 60
 DB 115 asasyeqyhslnelyswiefierhpdmltkihgssfekyplylylkysgkeqtaknal 174

QY 61 WTDCGTHAREWISPAFLMFIQH-----83
 Db 175 wldcgtharewlsapflcwfllghlclqfyglqylnllrvdlympvvnvdygyswkk 234
 QY 84 NEMARKNSRYANNHCIGTDLNRNFASKHWCCEGASSSSCSPTYCGLYPESPEPKAAVAS 143
 Db 235 nrmwknrsfyannhcltdlrrnfaskhwcceagassscsetyqglypesepekavas 294
 QY 144 FLRRNINIKAYISMHSYSHIYFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
 Db 295 flrrlnqikayismhsyqhlvfpystrskskdheelslvaseavraiektsknttryt 354
 QY 204 HGHSEETLYLAPGCGDDWYDLGIRYSFT 232
 Db 355 hghsetlylapggddwlydlylkyfst 383

RESULT 8
 AAB58129
 ID AAB58129 standard; Protein: 211 AA.
 AC AAB58129;
 DE 14-MAR-2001 (first entry)
 XX
 XX Lung cancer associated polypeptide sequence SEQ ID 467.
 DE
 XX Human: lung cancer associated protein; neuroprotective; cytostatic;
 KM cardioactive; immunomodulatory; muscular active; vulnery; cy
 KM gastrointestinal; nephrotoxic; antiinfective; gynecological;
 KM antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KM proliferative disorder; wound healing; infectious disease.
 XX
 OS Homo sapiens.
 PN WO20005180-A2.
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05918.
 PR 12-MAR-1999; 9905-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 PI Ruben SM.
 DR WPI; 2000-587514/55.
 DR N-PSDB; AAF18005.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 XX Claim 11; Page 944-945; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF1982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnery; gastrointestinal
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the

CC Identification and characterisation of the polynucleotide and protein
 CC sequences.
 CC
 SQ Sequence 211 AA;
 Query Match 61.0%; Score 816.5; DB 21; Length 211;
 Best Local Similarity 90.6%; Pred. No. 1,5e-78;
 Matches 154; Conservative 1; Mismatches 6; Indels 9; Gaps 2;

QY 71 WISPAFLMFI-----GHRMRKRNSRYANNHCIGTDLNRNFASKHWCCEGASSS 122
 Db 3 wis-mrlcmwmyunyswkknmrknsfyannhcltdlrrnfaskhwcceagasss 61
 QY 123 CSEYCGLYPESPEPEKAAVASFLRRNINIKAYISMHSYSHIYFPYSYTRSKSKDHEEL 182
 Db 62 csetyqglypesepekavasflrrlnqikayismhsyqhlvfpystrskskdheel 121
 QY 183 SLVASEAVRAIEKTSKNTRYTHGHSEETLYLAPGCGDDWYDLGIRYSFT 232
 Db 122 slvaseavraiektsknttrythghsetlylapggddwlydlylkyfst 171

RESULT 9
 AAG6562
 ID AAG6562 standard; Protein: 324 AA.
 AC AAG6562;
 DE 22-OCT-2001 (first entry)
 XX
 XX Bothops jararaca carboxypeptidase homologue protein.
 DE
 XX secreted metallocarboxypeptidase; immunomodulatory;
 KM antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
 KM osteopathic; antiarthritic; cerebroprotective; vasotropic; vulnery;
 KM antibacterial; antiviral; antitumoral; antiinflammatory; antidiabetic;
 KM antithematic; cytostatic; antiallergic; haemostatic; infection;
 cancer.
 XX
 OS Bothops jararaca.
 PN WO200157265-A1.
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001MO-US03783.
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 29-SEP-2000; 2000US-0676135.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Boyle BJ, Mize NK, Arterburn MC, Yeung G, Tang YT, Liu C;
 PI Dmanac RT, Wang M, Chen L, Yang Y;
 DR WPI; 2001-465705/50.
 XX
 PT Novel metallocarboxy peptidase-like polypeptide is useful in treating
 PT neurodegenerative, immune, etc. disorders -
 XX
 XX Example 3; Fig 1; 146pp; English.
 CC The invention relates to an isolated polypeptide comprising an amino
 CC acid sequence of 374, 41, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
 CC 354, 165, 374, 354, or 41 amino acids, fully defined in the
 CC specification, or its translated protein coding portion, mature protein
 CC coding portion, extracellular portion or active domain. The polypeptide
 CC is a human secreted metallocarboxypeptidase-like polypeptide. It
 CC is useful for treating mammalian diseases including:
 CC neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
 CC osteoporosis, osteoarthritis, disorders linked to reduced tissue

CC growth, Huntington's disease, amyotrophic lateral sclerosis, stroke, CC
CC reperfusion, wound healing, infectious diseases, immune and autoimmune
CC disorders such as severe combined immunodeficiency, rheumatoid
CC arthritis, Guillain-Barre syndrome, insulin dependent diabetes
CC mellitus, graft-versus-host disease and allergy, cancer, and
CC hemophilia. The polypeptide is also useful in medical imaging, as an
CC immunogen to generate antibodies, as a nutritional source or
CC supplement, and as a contraceptive. The present sequence is
CC a carboxypeptidase homologous protein of *Bothrops jararaca*.

50 Sequence 324 AA;

Query Match	45.88	Score 612.5	DB 22	Length 324
Best Local Similarity	45.08	Pred. No. 1.5e-56		
Matches 121, Conservative	44	Mismatches 63	Indels 41	Gaps 5

QY	2	SASV-YE0YHSJLNTYVSIEMIEFTRRHDDMLTKHIGSSFEKYEPLVYVKSQEDQATNAKI	60
		: : : : : : : : : : : : : : : : : : : : :	
Db	30	tagnyekaymswekaidwatadianempslvsrlqglttfegrrpmltkv-gkpvgnkkaal	88
QY	61	WIDCGIHAREMISPAFLCMTF-----GH	83
		: : : : : : : : : : : : : : : : : : : : : :	
Db	89	fidcgfharewispafocqwfreaavtrygketlmtqtlmkldfyflpblndgvyyswq	148
QY	84	NRMRRKNRSPFANNHCHGTDLNTRNFAKHMCEBAGSSSCSETPCGJLPSEPEPVKAVAS	143
		: : : : : : : : : : : : : : : : : : : : : :	
Db	149	srwrwtkctisvnaagctcigtldprnfida-wcsyvaasrnpsetcygskspsetckelad	207
QY	144	FLRRNRINOIKAYIMSHSYQHIIVPYVSYTTSRKSQDHELISIVASEAVRAIEKTSKNTRY	203
		: : : : : : : : : : : : : : : : : : : : :	
Db	208	ftrnrslqgltyllhaysgmlllypyrytydltensnkklnslakealrel-kvlfgreyt	266
QY	204	HGHSSEITLYLAPGGDDMIYDLGIKTISFT	232
		: : : : : : : : : : : : : : : : : : : : :	
Db	267	ygpapaatlypaagsddwayoqgiklyaf	295

RESULT 10
AAW00602

ID AAW00602 standard; Protein; 307 AA.

AC AAW00602;

DT 09-NOV-1996 (first entry)

DE Rat carboxypeptidase B mature enzyme.

KW Carboxypeptidase-B; pro-enzyme; protease; insulin.
...

OS Rattus sp.

PN W09623064-A1.

PD 01-AUG-1996.

PF 25-JAN-1996; 96WO-US00995.

PR 25-JAN-1995; 95US-0378233.

PA (BIOT-) BIOTECHNOLOGY GEN CORP.

PI Fulga N, Gorecki M, Hartman J, Mendelovitch S;
xx

DR WPI; 1996-362688/36.

XX DE 041507

PT Purified active recombinant carboxypeptidase B prodn. - by
PT expressing DNA encoding the pro-enzyme, folding and enzymatic
PT cleavage to give active form. used e.g. for insulin prodn.

PS Example 1; Page 37-38; 49pp; English.

CC The amino acid sequence of the rat mature carboxypeptidase-B
CC (CPB) (AA000602) was deduced from a cDNA sequence (see also AA1735760).
CC The sequence of the activation peptide (AA000601) of pro-CPB has
CC been deduced. Pro-CPB (Pref with a C290S mutation) can be
CC expressed in host cells, e.g. *Escherichia coli*, and subsequently
CC recovered, refolded and cleaved with trypsin to yield the active
CC enzyme. CPB produced this way is cheaper than porcine pancreatic
CC enzyme, and is free of other proteases.

SQ Sequence 307 AA;

Query Match	45.68;	Score 610;	DB 17;	Length 307;
Best Local Similarity	45.08;	Pred. No. 2.6e-56;		
Matches 121; Conservative	43;	Mismatches 65;	Indels 40;	Gaps 5

```

OY      1 ASASYEDDYHSILNELYSMIEETITERRPMPLKIHIGSFEKYPPLVYLKASGEOTAKNAI 60
Db      1 asghayctkynmwetleawidqvatcdnpdlvtqsvigtlttegrmyviki -gktprkpaI 59
OY      61 WIDCGIHAREMISPACFLMFI-----GH----- 83
Db      60 fidecgfharewispacfcwifarevrltyngelhmkgjldeldfyvlrvpvnidgyvltwk 119
OY      84 NRMWKNKNSFFVANNHCIGTDLNRMPASKHMCBEGASSCCSEYTCGLYLPSEPRKVAS 143
Db      120 dirmwktktstmasgcljvcdpnrlf -nagweaveaatspsceetycgspapekecktaIad 178
OY      144 FLRRNINQIKAYISMHSYQIHVEPYYSYTRSKSDHEELSLVASAEVRAIEKTSKNRRT 203
Db      179 firmIstckaylcthsygsqmmlylpyssdyvlpenyeelnaIvkaekel -atlhgkyt 237
OY      204 HGHSSETLYLAPGGDDMIYLDGKISFT 232
Db      238 yypgacltlylpaagsddwsydgqIkystf 266

```

RESULT 11
AAR75132

ID AAR75132 standard; Protein; 306 AA

AC AAR75132;

DT 13-MAR-1996 (first entry)

DE Porcine carboxypeptidase B.

KW Procarboxypeptidase B; carboxypeptidase B; Pichia; PCPB;

КМ ПРОСБВ.

OS Sus scrofa.

PN W09514096-A1

PD 26-MAY-1995.

PF 16-NOV-1994; 94WO-US13142.

PR 16-NOV-1993; 93US-015

PA (ELL) LILLY & CO ELL.

PI Fayerman JT, Greener

XX

DR N-PSDB; AAQ90601.

PT DNA encoding porcine carboxypeptidase B - used for transforming host cells, partic. *Pichia* species, for prodn. of the enzyme

PS Example 1; Page 18; 34pp; English.
XX

The porcine carboxypeptidase B coding sequence can be placed in a bacterial or pre-f. Pichia yeast expression vector. The expression vector further comprises the signal peptide of either human serum albumin (designated pLGD3 - NRRL B-21029), premating factor alpha (designated pF4489 - NRRL B-21028) or the porcine proCB signal peptide (pF474 - NRRL B-21032) or the porcine proCB signal peptide (designated pLGD27 - NRRL B-21027). The method can be used for producing large amounts of porcine carboxypeptidase B and when produced in Pichia yeast, the protein does not need solidification or folding. The produced enzyme is then used for pre-f. cleaving basic residues from the carboxyterminus of proteins.

SQ Sequence 306 AA;

Query Match	43.5%	Score 582;	DB 16;	Length 306;
Best Local Similarity	44.3%	Pred. NC. 2.5e-53;		
Matches 117;	Conservative 40;	Mismatches 67;	Indels 40;	Gaps 5;

QY	6	YEOHSHSINLEYSWIEFTTERPDMLTIIHGSSPEKPLVILKVGSGEOTAKNAIWIIDCG	65
Db	6	yekymwetleawxkygtscnpdlstaltagtltlgnmilylkv-gkxppnukpaimdcg	64
QY	66	IHAREWISPAFLMFT-----GH-----NEMWR	88
Db	65	fharewshahicqfwfreaavlytyeshaueflnkldfyvlpvlnldqlywtctkmwr	124
QY	89	KNSFVYNNHNCIGDILRNFSKRWCEGASSSSCSTGYGLPESPEPKAVASFARN	148
Db	125	ktsrnaqtcicidpnnrl-dagwcttcgaastcdpdcdecygsaaaseketaladifrm	183
QY	149	INOIKAVISMHSYSOHIVPEFYSYTRSKSKOHEELSLVAESAVALAEKTSKNTRYTHGGS	208
Db	184	lsaikeylthysgmilypysdykripemnaeilmnlakaavel-atlygtvtytgyga	242
QY	209	EFLYLAPGGDDMIYDGLKXYSFT	232
Db	243	tllypaagddwaydgglyastl	266

RESULT	12
AAU04477	
ID	AAU04477
XX	standard; Protein; 306 AA
AC	AAU04477;

DT 26-SEP-2001 (first entry)
XX
DE Porcine carboxypeptidase B (Cpb) protein.
DE

carboxypeptidase B; CPB; p4g: protease activity; trypsin; chymotrypsin; animal product; purification; contaminant; biopharmaceutical agent; zinc-containing pancreatic exopeptidase; insulin; proteolytic processing; protein sequence analysis.

Sus scrofa.

PN WO200151624-A2.

PD 19-JUL-2001.

12-JAN-2001; 2001WO-US000008.

12-JAN-2000; 2000US-0175781.

PA (ELIL) LILLY & CO ELI.

Hanquiter JM;

WPI; 2001-442143/47.

PT New carboxypeptidase B, which is free of animal products and
PT contaminating enzyme activity, useful for producing biopharmaceutical
PT agents (e.g. Insulin) and biologically active polypeptides, as well as
PT in protein sequence analysis

PS Claim 4; Page 41-42; 42pp; English.
XY

The sequence represents porcine carboxypeptidase B (CPB), which is essentially free of contaminating protease activity (e.g. trypsin or chymotrypsin activity) and free of animal products, bacterial cell components and other contaminants that result from purification processes. CPB is a zinc-containing pancreatic exopeptidase which specifically removes C-terminal arginine, lysine and ornithine from peptides. The CPB of the invention is useful for commercial and research purposes, e.g. in production of biopharmaceutical agents, such as insulin and other biologically active polypeptides, as well as in protein sequence analysis. Compositions containing CPB are useful in preparing proteins which are in need of proteolytic processing in order to form the mature protein.

sq **Sequence** **306 AA;**

Query Match	43.58;	Score 582;	DB 22;	Length 306;
Best Local Similarity	44.38;	Pred. No. 2.5e-53;		
Matches 117; Conservative	40;	Mismatches 67;	Indels 40;	Gaps 5

[illegible]

RESULT	13
ID	AAR75131
NAME	standard; Protein; 404 AA.

AC AAR75131;
YY

DT 13-MAR-1996 (first entry)
 VV

Porcine Tyr-His-Met Procarboxypeptidase B.

Procarboxypeptidase B; carboxypeptidase B; Pichia; PCPB; human serum albumin.

PROCBB.
KW

Sus scrofa

PN W09514096-A1.

PD 26-MAY-1995.

16-NOV-1994;

PR 16-NOV-1993;
XX

PA (ELIL) LILLY & CO ELI

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:49:06 ; Search time 55.22 Seconds

(without alignments)
770.676 Million cell updates/sec

Title: US-09-980-881-4

Perfect score: 1338

Sequence: 1 ASASYEQYHSLNEITSMIE.....IKTSFSPNPVEKLPLSLK 246

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database : SPTREMBL_19:**
1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mhc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp Vertebrate:**
14: sp Unclassified:**
15: sp_virus:**
16: sp_bacteriaph:**
17: sp_Archeap:**

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	1327	99.2	360	4	Q9P2Y6
2	1239.5	92.6	423	4	Q15114
3	1234.5	92.3	423	4	Q961Y4
4	1105.5	82.6	422	11	Q9JH86
5	1095.5	81.9	422	11	Q9EQV9
6	1095.5	81.9	422	11	Q9EQV9
7	612.5	45.8	416	13	Q9PUF2
8	582	43.5	416	6	Q9XSP3
9	567	42.4	417	4	Q96B08
10	557	41.6	412	11	P97397
11	552	41.3	417	4	Q96E94
12	537	40.1	279	11	Q9CVD1
13	488.5	35.5	419	6	Q9TV85
14	477.5	35.7	247	4	Q96K29
15	465.5	34.8	419	4	Q9BS67
16	459	34.3	417	4	Q96ON3

17	459	34.3	417	4	Q96A12	Q96A12 homo sapien
18	456	34.1	198	4	Q9NT18	Q9NT18 homo sapien
19	426	31.8	1192	5	Q9W475	Q9W475 drosophila
20	424	31.7	424	5	Q9VLE87	Q9VLE87 drosophila
21	419	31.3	422	5	Q9VLE82	Q9VLE82 drosophila
22	419	31.3	440	5	Q9W478	Q9W478 drosophila
23	403	30.1	323	5	Q19121	Q19121 caenorhabd
24	396.5	29.6	584	5	Q9R2H1	Q9R2H1 caenorhabd
25	385.5	28.8	453	5	Q9V342	Q9V342 drosophila
26	382.5	28.6	424	5	Q9VS66	Q9VS66 drosophila
27	380	28.4	430	5	Q9VLE86	Q9VLE86 drosophila
28	377.5	28.2	423	5	Q61532	Q61532 drosophila
29	375	28.0	427	5	Q9U9K2	Q9U9K2 aedes aegy
30	370.5	27.7	419	5	Q9VLE21	Q9VLE21 drosophila
31	366.5	27.4	455	5	Q9R2C6	Q9R2C6 caenorhabd
32	357.5	26.7	400	5	Q9VSM7	Q9VSM7 drosophila
33	355	26.5	467	5	Q9VCM8	Q9VCM8 drosophila
34	354	26.5	581	5	Q23318	Q23318 caenorhabd
35	349	26.1	415	5	Q9VS64	Q9VS64 drosophila
36	344.5	25.7	415	5	Q9VS68	Q9VS68 drosophila
37	344.5	25.7	418	5	Q9VLEJ8	Q9VLEJ8 drosophila
38	344.5	25.7	540	5	Q9R2H2	Q9R2H2 caenorhabd
39	342	25.6	665	5	Q9XU75	Q9XU75 caenorhabd
40	340	25.4	1430	5	Q9XU86	Q9XU86 drosophila
41	336	25.1	545	5	Q9N3S6	Q9N3S6 caenorhabd
42	331.5	24.8	161	13	Q90W12	Q90W12 brachydanio
43	325.5	24.3	312	5	Q9VRZ3	Q9VRZ3 drosophila
44	319.5	23.9	173	11	Q9D961	Q9D961 mus musculu
45	302	22.6	344	5	Q9VS65	Q9VS65 drosophila

ALIGNMENTS

RESULT 1
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B-LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid-9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92042093; PubMed-1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL: AB011969; BAA90475.1; -.
DR HSSP: P00730; 2CTC.
DR MEROPS: M14.009; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept; 2.
DR PRINTS: PR00765; CRBOXYPTASEA.
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

Query Match 99.2%; Score 1327; DB 4; Length 360;
Best Local Similarity 99.2%; Pred. No. 6,9e-121;
Matches 244; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 ASASYEQYHSLNLEYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKSGKEQTAKNAI 60
DB 115 ASASYEQYHSLNLEYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKSGKEQTAKNAI 174
QY 61 WIDCGIHAREWISPAFCFLMFIQHNHMRKRNSFYANNHCIGDNLNRFASKWCEGAS 120
DB 175 WIDCGIHAREWISPAFCFLMFIQHNHMRKRNSFYANNHCIGDNLNRFASKWCEGAS 234
QY 121 SSCSEYCGLYPESEPEVAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 180
DB 235 SSCSEYCGLYPESEPEVAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 294
QY 181 ELISVAEAVRAIEKTSKTRTHGSGETLYLAPGGDDWIYDLGIKYSFNSPVEKL 240
DB 295 ELISVAEAVRAIEKTSKTRTHGSGETLYLAPGGDDWIYDLGIKYSFNSPVEKL 354
QY 241 LPLSLK 246
DB 355 LPLSLK 360

```

```

RESULT 2
Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PCPB PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE-20042093; PubMed-1939207;
RA Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Dayana D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL J. Biol. Chem. 266:21833-21838 (1991).
DR EMBL; M75106; AAA60042.1; -
DR HSSP; P00730; 1PPT.
DR MEROPS; M14.009; -
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
FT CHAIN 23
SQ SEQUENCE 423 AA; 48442 MW; 9B383272FE6E79F4 CRC64;

```

```

Query Match 92.6%; Score 1239.5; DB 4; Length 423;
Best Local Similarity 86.2%; Pred. No. 2.8e-112;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```

```

QY 1 ASASYEQYHSLNLEYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKSGKEQTAKNAI 60
DB 115 ASASYEQYHSLNLEYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKSGKEQTAKNAI 174
QY 61 WIDCGIHAREWISPAFCFLMFIQHNHMRKRNSFYANNHCIGDNLNRFASKWCEGAS 120
DB 175 WIDCGIHAREWISPAFCFLMFIQHNHMRKRNSFYANNHCIGDNLNRFASKWCEGAS 234
QY 121 SSCSEYCGLYPESEPEVAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 180
DB 235 SSCSEYCGLYPESEPEVAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 294
QY 181 ELISVAEAVRAIEKTSKTRTHGSGETLYLAPGGDDWIYDLGIKYSFNSPVEKL 240
DB 295 ELISVAEAVRAIEKTSKTRTHGSGETLYLAPGGDDWIYDLGIKYSFNSPVEKL 354
QY 241 LPLSLK 246
DB 355 LPLSLK 360

```

```

DB 295 FLRRNINQIKAYISMHSYQHIYFPYSYRSKDHLEISLVAEAVRAIEKTSKTRTY 354
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 355 HGHGSETLYLAPGGDDWIYDLGIKYSFT 383

```

```

RESULT 3
Q0961Y4 PRELIMINARY; PRT; 423 AA.
AC Q0961Y4
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B2 (PLASMA).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKLETAL MUSCLE.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007057; AA07057.1; -
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6E7CF5 CRC64;

```

```

Query Match 92.3%; Score 1234.5; DB 4; Length 423;
Best Local Similarity 85.9%; Pred. No. 8.5e-112;
Matches 231; Conservative 0; Mismatches 1; Indels 37; Gaps 1;

```

```

QY 1 ASASYEQYHSLNLEYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKSGKEQTAKNAI 60
DB 115 ASASYEQYHSLNLEYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKSGKEQTAKNAI 174
QY 61 WIDCGIHAREWISPAFCFLMFIQHNHMRKRNSFYANNHCIGDNLNRFASKWCEGAS 120
DB 175 WIDCGIHAREWISPAFCFLMFIQHNHMRKRNSFYANNHCIGDNLNRFASKWCEGAS 234
QY 121 SSCSEYCGLYPESEPEVAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 180
DB 235 SSCSEYCGLYPESEPEVAVASFLRRNINQIKAYISMHSYQHIYFPYSYRSKDH 294
QY 181 ELISVAEAVRAIEKTSKTRTHGSGETLYLAPGGDDWIYDLGIKYSFNSPVEKL 240
DB 295 ELISVAEAVRAIEKTSKTRTHGSGETLYLAPGGDDWIYDLGIKYSFNSPVEKL 354
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 355 HGHGSETLYLAPGGDDWIYDLGIKYSFT 383

```

```

RESULT 4
Q09JH6 PRELIMINARY; PRT; 422 AA.
AC Q09JH6
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE R (THROMBIN-ACTIVATABLE FIBRINOLYSIS INHIBITOR)
DE ((1110032P04RIK PROTEIN)).
GN CPB2 OR TAFT OR 1110032P04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-20341711; PubMed-10876383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,

```

RT whereas carboxypeptidase N is not.;"

RL J. Immunol. 165:1053-1058(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-20201996; PubMed-10739389;

RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,

RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.,

RT "Characterization of mouse thrombin-activatable fibrinolysis

RT Inhibitor.;"

RL Thromb. Haemost. 83:297-303(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-C57BL/6J; TISSUE-EMBRYO;

RX MEDLINE-21085660; PubMed-11217851;

RA Kawai J., Shitagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,

RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,

RA Felschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schiml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmung L.,

RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.;"

RL Nature 409:685-690(2001).

DR EMBL; AB021968; BAB3402.1; -

DR EMBL; AF164524; AAF62385.1; -

DR EMBL; AK004045; BAB2141.1; -

DR HSSP; P00730; 2CTC.

DR MEROPS; M14.009; -

DR MGD; MGI:1891837; Cpb2.

DR InterPro; IPR003146; Propep_M14.

DR InterPro; IPR000834; Zn_carboxypept.

DR Pfam; PF02244; Propep_M14; 1.

DR Pfam; PF00246; Zn_carboxypept. 1.

DR PRINTS; PR00765; CRBOXYPTASEA.

DR SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

Query Match 82.6%; Score 1105.5; DB 11; Length 422;

Best Local Similarity 77.0%; Pred. No. 3.1e-99;

Matches 207; Conservative 10; Mismatches 15; Indels 37; Gaps 1;

QY 1 ASASYEQYHSLNEIYSWIFETTERHPDMLTKIHGSEFEKYPPLYLVKSGKEQTAKNAI 60

DB 114 ASASYEQYHSLNEIYSWIFETTERHPDMLTKIHGSEFEKYPPLYLVKSGKEQTAKNAI 173

QY 61 WIDCGIHAREMISPAFCLEWFGH----- 83

DB 174 WIDCGIHAREMISPAFCLEWFGH----- 233

QY 84 NRMWRKRNSFYANNHCIGTDLNRNFASKHWCEGASSSCSEYCGLYPSEPEVKAVAS 143

DB 234 NRMWRKRNSFYANNHCIGTDLNRNFASKHWCEGASSSCSEYCGLYPSEPEVKAVAS 293

QY 144 FLRRNINQIKAYISMHSYSHIYFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203

DB 294 FLRRNINQIKAYISMHSYSHIYFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 353

QY 204 HGHGSETLYLAPGGDDWIDYDLGIKYSFT 232

DB 354 HGHGSETLYLAPGGDDWIDYDLGIKYSFT 382

RESULT 5

ID Q9EQV9 PRELIMINARY; PRT; 422 AA.

AC Q9EQV9;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update).

DE PRE-PROCARBOXYPEPTIDASE R.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID-10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-204711387; PubMed-11021404;

RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,

RA Okada N., Okada H.;

RT "Molecular cloning and partial characterization of rat

RT procarboxypeptidase R and carboxypeptidase N.;"

RL Microbiol. Immunol. 44:719-728(2000).

DR HSSP; P00730; 2CTC.

DR InterPro; IPR003146; Propep_M14.

DR InterPro; IPR000834; Zn_carboxypept.

DR Pfam; PF02244; Propep_M14; 1.

DR Pfam; PF00246; Zn_carboxypept. 1.

DR PRINTS; PR00765; CRBOXYPTASEA.

KW Carboxypeptidase.

SO SEQUENCE 422 AA; 48826 MW; FFFD32A51A9366C8 CRC64;

Query Match 81.9%; Score 1095.5; DB 11; Length 422;

Best Local Similarity 75.8%; Pred. No. 2.9e-98;

Matches 204; Conservative 11; Mismatches 17; Indels 37; Gaps 1;

QY 1 ASASYEQYHSLNEIYSWIFETTERHPDMLTKIHGSEFEKYPPLYLVKSGKEQTAKNAI 60

DB 114 ASASYEQYHSLNEIYSWIFETTERHPDMLTKIHGSEFEKYPPLYLVKSGKEQTAKNAI 173

QY 61 WIDCGIHAREMISPAFCLEWFGH----- 83

DB 174 WIDCGIHAREMISPAFCLEWFGH----- 233

QY 84 NRMWRKRNSFYANNHCIGTDLNRNFASKHWCEGASSSCSEYCGLYPSEPEVKAVAS 143

DB 234 NRMWRKRNSFYANNHCIGTDLNRNFASKHWCEGASSSCSEYCGLYPSEPEVKAVAS 293

QY 144 FLRRNINQIKAYISMHSYSHIYFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203

DB 294 FLRRNINQIKAYISMHSYSHIYFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 353

QY 204 HGHGSETLYLAPGGDDWIDYDLGIKYSFT 232

DB 354 HGHGSETLYLAPGGDDWIDYDLGIKYSFT 382

RESULT 6

ID Q9QZFO PRELIMINARY; PRT; 422 AA.

AC Q9QZFO;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE CARBOXYPEPTIDASE U.

GN CPB2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID-10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-B6/CHAPLJ;

RA He Y.C., Broze G.;

RT "Isolation and characterization of mouse liver carboxypeptidase B
 RT gene.";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A181818; AAF00528.1; -
 DR HSSP: P00730; 5CPA.
 DR MEROPS: M14.009; -
 DR MGD: MGI:1891837; Cpb2.
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxpept.
 DR Pfam: PF02244; Zn_carboxpept.
 DR Pfam: PF00246; Zn_carboxpept; 1.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 SQ SEQUENCE 422 AA; 48832 MW; B53FFB09943954E CRC64;

Query Match 81.9%; Score 1095.5; DB 11; Length 422;
 Best Local Similarity 76.2%; Pred. No. 2.9e-98;
 Matches 205; Conservative 11; Mismatches 16; Indels 37; Gaps 1;

OY 1 SASASYEYVHSLNETYSWIEFTTERHPDMLTKIHGSSFEKPYLVKSGKEQTAKNAI 60
 DB 114 SASASYEYVHSLNETYSWIEFTTERHPDMLTKIHGSSFEKPYLVKSGKEQTAKNAI 173
 OY 61 WIDCGIHAREWISPAFCLEMTFI----- 83
 DB 174 WIDCGIHAREWISPAFCLEMTFI----- 83
 OY 84 NMMWRKNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVAS 143
 DB 234 NMMWRKNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVAD 293
 OY 144 FLRRNINQIKAYISMHSYQHIPEYSTRSKSDHELSLVASAVAIETSKNRYT 203
 DB 234 FLRRNINQIKAYISMHSYQHIPEYSTRSKSDHELSLVASAVAIETSKNRYT 353
 OY 204 HGHGSETLYLAPGGGDDMIYDLGIRYSFT 232
 DB 354 HGHGSETLYLAPGGGDDMIYDLGIRYSFT 382

RESULT 7
 OYPUF2 PRELIMINARY; PRT; 416 AA.
 AC OYPUF2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 13, Last annotation update)
 DE CARBOXYPEPTIDASE HOMOLOG.
 OS Bothrops jararaca (Jararaca).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
 OC Viperidae; Crotalinae; Bothrops.
 OX NCBI_TaxID=8724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
 RT "Screening of Bothrops jararaca pancreas cDNA library."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A190274; AAF01344.1; -
 DR HSSP: P09955; 1NSA.
 DR MEROPS: M14.003; -
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxpept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxpept; 1.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 SQ SEQUENCE 416 AA; 47723 MW; 4F999854DD72B7A7 CRC64;

Query Match 45.8%; Score 612.5; DB 13; Length 416;
 Best Local Similarity 45.0%; Pred. No. 2.9e-51;

Matches 121; Conservative 44; Mismatches 63; Indels 41; Gaps 5;
 OY 2 SASY-YEYVHSLNETYSWIEFTTERHPDMLTKIHGSSFEKPYLVKSGKEQTAKNAI 60
 DB 111 TAGYTERKINSWEKIDAMTADIANENPSLVSLQIGTFEGRPMLKLV-GKPGYNKNAI 169
 OY 61 WIDCGIHAREWISPAFCLEMTFI-----GH 83
 DB 170 FIDCGFHAREWISPAFCQMFVEAVRTYGTETIMQTLNKIDFYILPLVNLIDGYVYSNKQ 229
 OY 84 NMMWRKNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVAS 143
 DB 230 NMMWRKNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVAD 288
 OY 144 FLRRNINQIKAYISMHSYQHIPEYSTRSKSDHELSLVASAVAIETSKNRYT 203
 DB 269 FLRRNINQIKAYISMHSYQHIPEYSTRSKSDHELSLVASAVAIETSKNRYT 347
 OY 204 HGHGSETLYLAPGGGDDMIYDLGIRYSFT 232
 DB 348 YGPGATYTPAAGSGDDMIYDLGIRYSFT 376

RESULT 8
 OYXSP3 PRELIMINARY; PRT; 416 AA.
 AC OYXSP3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PROCARBOXYPEPTIDASE B PRECURSOR (EC 3.4.17.2).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PANCREAS;
 RX MEDLINE=99321926; PubMed=10391940;
 RA Ventura S., Villegas V., Sturner J., Larson J., Vendrell J.,
 RS Hersberger C., Aviles F.;
 RT "Mapping the Pro-region of carboxypeptidase B by protein engineering,
 RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
 RL J. Biol. Chem. 274:11925-11933(1999).
 DR EMBL: AJ13775; CAB46991.1; -
 DR HSSP: P09955; 1PBA.
 DR MEROPS: M14.003; -
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxpept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxpept; 1.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 DR Signal: Carboxypeptidase; Hydrolase.
 FT STGNL 1 15 POTENTIAL.
 FT CHAIN 16 416 PROCARBOXYPEPTIDASE B.
 SQ SEQUENCE 416 AA; 47381 MW; 5FAVFCED7B45AE6A CRC64;

Query Match 43.5%; Score 582; DB 6; Length 416;
 Best Local Similarity 44.3%; Pred. No. 2.7e-48;
 Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;
 OY 6 YEYVHSLNETYSWIEFTTERHPDMLTKIHGSSFEKPYLVKSGKEQTAKNAIIDDG 65
 DB 116 YEKYNNMETIEAWTKQVTSNPDLISRTAIGTFGLNNIYLLKV-GKPPNKPALFMDCG 174
 OY 66 IHAREWISPAFCLEMTFI-----GH-----NRMWR 88
 DB 175 FHAREWISPAFCQMFVREAVLTGYESHMTFELNKIDFYILPLVNLIDGYIYTWTKRMWR 234
 OY 89 KNSFYANNHCTGDLNRNFAKHWCEGASSSCSEYCYGLYPESEPEYKAVASFLRRN 148

DB 235 KTRSTNAGTCTCTGDPNRRF-DAGMCTGASTDPDCEITCGSAASESEKTKALADPIRNN 293
 OY 149 INQIKAYISMHSYQSHIVFPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
 DB 294 ISSIKAYILHSYSQMIIPYSYDVKLPENNAELNLAKAAVKEL-ATLYGTRKTYTGPGA 352
 OY 209 ETLYLAPGGDDMIYDLGIKYSFT 232
 DB 353 TTIYPAGGSDMAYDQGIKYSFT 376

RESULT 9
 O96B08 PRELIMINARY: PRT: 417 AA.
 AC O96B08:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYDROTICAL 47.4 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ADRENAL CARCINOMA;
 RA Strausberg R.;
 DR Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC015338; AAH15338.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 417 AA; 47367 MW; EBBB98B27F5D5AF9 CRC64;

Query Match 42.4%; Score 567; DB 4; Length 417;
 Best Local Similarity 41.3%; Pred. No. 7.7e-47;
 Matches 113; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

OY 1 ASASYEYHSLNIEYWIETTERHPDMLTKIHGSEFEKYPILYLVKSGEQAKNAI 60
 DB 111 ATGHSYKKNKMETEATQVATENPALISRSVIGTFEGRAIYLKV-GRAGQKPAI 169
 OY 61 WIDCGIHAREWISPAFCLEFI-----GH----- 83
 DB 170 FMDGCFHAEWISPAFCQWFREAVRTYGTREIQTVELDKLDFYLPVINDIGYITWTK 229
 OY 84 NEMWKRNSFYANNHCIGTDLNRRNFASKHMCCEGASSSCSEYTCGLYPESEPEYKAVAS 143
 DB 230 SFPMKTRSTHGGSSCIGTDPNRRF-DAGMCEIGASRNCDEITCGPAESEKTKALAD 288
 OY 144 FLRRINQIKAYISMHSYQSHIVFPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYT 203
 DB 289 FIRNKLSSIKAYILHSYSQMIIPYSYAVKLGENNAELNLAKAVKEL-ASLHGTRKT 347
 OY 204 HGHSEETLYLAPGGDDMIYDLGIKYSFT 232
 DB 348 YGPGATTIYPAGGSDMAYDQGIKYSFT 376

RESULT 10
 P97597 PRELIMINARY: PRT: 412 AA.
 AC P97597:
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MAST CELL CARBOXYPEPTIDASE A PRECURSOR (FRAGMENT).
 GN R-CPA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE DAWLEY;
 RX MEDLINE-97149430; PubMed-8996238;
 RA Lutzelschab C., Fejler G., Aveskogh M., Hellman L.;
 RT "Secretory granule proteases in rat mast cells. Cloning of 10
 RT different serine proteases and a carboxypeptidase A from various rat
 RT mast cell populations." (1997).
 RL J. Exp. Med. 185:13-29(1997).
 DR EMBL: U67914; AAB48267.1; -
 DR HSSP: P09955; INSA.
 DR MEROPS: M14.010; -
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_cathopept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_cathopept; 1.
 DR PRINTS: PR00765; CARBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Signal; Carboxypeptidase.
 FT NON_TER 1
 FT SIGNAL <1 10 POTENTIAL.
 FT CHAIN 11 412 MAST CELL CARBOXYPEPTIDASE A.
 SQ SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

Query Match 41.6%; Score 557; DB 11; Length 412;
 Best Local Similarity 43.0%; Pred. No. 7.1e-46;
 Matches 113; Conservative 38; Mismatches 72; Indels 40; Gaps 5;

OY 6 YRDYHSLNIEYWIETTERHPDMLTKIHGSEFEKYPILYLVKSGEQAKNAIIDDG 65
 DB 111 YAKYNDNMKIVSTKEMVKEHPKEMVSRIRKISTVEDNPPLYVKI-GRKGERRAIPMDG 169
 OY 66 IHAREWISPAFCLEFI-----GHN-----RMR 88
 DB 170 IHAREWISPAFCQWFYQAAKSGKKNKTKLLDRNPFYLPVFNVDGYIMSTKRRMR 229
 OY 89 KNSFYANNHCIGTDLNRRNFASKHMCCEGASSSCSEYTCGLYPESEPEYKAVASFLRN 148
 DB 230 KNSKRNPNSTCIGTDLNRRF-DVSMDSPTNDPCLSYRGPAESEKTKAVTNFRSH 288
 OY 149 INQIKAYISMHSYQSHIVFPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
 DB 289 LNSIKAYITFHYSQMLPEYGTITKLPNHODLVKARIADVL-SSRETRYIGPIA 347
 OY 209 ETLYLAPGGDDMIYDLGIKYSFT 231
 DB 348 STTYKTSGLDMAYDQGIKHTF 370

RESULT 11
 O96E94 PRELIMINARY: PRT: 417 AA.
 AC O96E94:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE SIMILAR TO CARBOXYPEPTIDASE A3 (MAST CELL).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BONE MARROW, AND ACUTE MYELOCYTIC LEUKEMIA;
 RA Strausberg R.;
 DR Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC012613; AAH12613.1; -
 KW Carboxypeptidase.
 SQ SEQUENCE 417 AA; 48714 MW; F781EEF23F570E98 CRC64;

Query Match 41.3%; Score 552; DB 4; Length 417;

DB 294 NQGN-1KAFISHSYQLLIPYGYKTEPNAPKDELQISKAVAL-TSLYGTKEFG 351
 QY 206 HGSFTLYLAPGGDDMIYDLGIRYSFT 232
 DB 352 SITTTIYQASGTTDWTYNGIKYSFS 378

RESULT 14
 Q96KZ9 PRELIMINARY; PRT; 247 AA.

AC 096KZ9;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PANCREAS;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC015104; AF015104.1; -.
 SQ Hypothetical protein.
 KW
 SEQUENCE 247 AA; 28181 MW; 914DD8BA162C1082 CRC64;

Query Match 35.7%; Score 477.5; DB 4; Length 247;
 Best Local Similarity 43.8%; Pred. No. 2e-38;
 Matches 91; Conservative 30; Mismatches 48; Indels 39; Gaps 4;

QY 62 IDGIIHAREWISPAFLMFI-----GH-----N 84
 DB 1 MCGFHAREWISPAFCOMFVREAVRTYGRIOYTELLDKIDFVPLNLNDGIYITWTKS 60
 QY 85 RMKRKRSTYANNHCIGTDLNRNFASKHWCCEGASSSCSEYCGLYPESEPEYKAVASF 144
 DB 61 RFWKRRTSTHTGSCICIGTPNRF-DAGWCEIGASRNPCEYCGPAESEKETKALADF 119
 QY 145 LRNNIOIKAYISMHSYQHLVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTH 204
 DB 120 IRKKLSIRAYLTISHISQMLTPYSAVKLGNNALNALAKATYKEL-ASLHGITYTY 178
 QY 205 HGSFTLYLAPGGDDMIYDLGIRYSFT 232
 DB 179 GPATTIYPAAGSGDDMAYDQIRYSFT 206

RESULT 15
 Q9BS67 PRELIMINARY; PRT; 419 AA.
 AC 09BS67;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CARBOXYPEPTIDASE A1 (PANCREATIC).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PANCREAS;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC005279; AF05279.1; -.
 DR HSSP; P00730; 1PT.
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carlopept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carlopept; 1.

DR PRINTS; PR00765; CARBOXYPTASE.
 DR PROSITE; PS00132; CARBOXYPEPT_2N_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_2N_2; 1.
 SQ SEQUENCE 419 AA; 47159 MW; 7855759PAEF982DD CRC64;

Query Match 34.8%; Score 465.5; DB 4; Length 419;
 Best Local Similarity 37.9%; Pred. No. 5.9e-37;
 Matches 100; Conservative 44; Mismatches 79; Indels 41; Gaps 5;

QY 6 YEQYHSINELYSWIEFTTERHPDMLTKIHGSSPEKPLVLYKSGKEQTAKNAIWDGC 65
 DB 119 YATYHTEELYDFLDLVARNPHLVSKIQIGNYEGRPYVLKS-TGSKRPPIWIDTG 177
 QY 66 IHAREWISPAFLMFI-----GHNMNR 88
 DB 178 IHSHEWYQASGVFAKKITQDYGDAFTALDITDIFLEYTNPDGFAFTHSTNMMR 237
 QY 89 KNRSEYANNHCIGTDLNRNFASKHWCCEGASSSCSEYCGLYPESEPEYKAVASF 148
 DB 238 KTRSHTAGSLCIGVDPNRNDAGFGL-SGASSNPSCSEYGRKGFANSEVEVKSIVDFVKD 296
 QY 149 INQIKATISMHSYQHLVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHS 208
 DB 297 GN-1KAFISHSYQLLIPYGYKTEPNAPKDELQISKAVAL-ASLYGTKEFGSII 354
 QY 209 ETLYLAPGGDDMIYDLGIRYSFT 232
 DB 355 KATYQASGTTDWTYSGIKYSFT 378

Search completed: September 18, 2002, 16:49:07
 Job time: 299 sec

This Page Blank (u.s.)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:46:03 ; Search time 28.95 Seconds

(without alignments)
207.554 Million cell updates/sec

Title: US-09-980-881-4

Perfect score: 1338

Sequence: 1 ASASYEYQYHSLNEIYSWIE.....IKTSFTSNPPVKLPLSLK 246

Scoring table: BLOSUM62

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA.*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*

2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*

3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*

4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*

5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep.*

6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239.5	92.6	423	1	US-07-649-591B-3
2	1239.5	92.6	423	1	US-08-277-540-3
3	1239.5	92.6	423	1	US-08-430-787A-3
4	1239.5	92.6	423	2	US-08-869-057-2
5	610	45.6	307	2	US-08-782-760-6
6	610	45.6	307	5	PCT-US96-00995-6
7	582	43.5	306	1	US-08-696-139-4
8	582	43.5	404	1	US-08-696-139-2
9	567	42.4	415	2	US-08-860-882A-57
10	561	41.9	417	1	US-07-649-591B-7
11	561	41.9	417	1	US-08-277-540-7
12	561	41.9	417	1	US-08-430-787A-7
13	554	41.4	396	1	US-07-649-591B-4
14	554	41.4	396	1	US-08-277-540-4
15	554	41.4	396	1	US-08-430-787A-4
16	552	41.3	417	1	US-07-649-591B-6
17	552	41.3	417	1	US-08-277-540-6
18	552	41.3	417	1	US-08-430-787A-6
19	549	41.0	613	4	US-09-171-945-113
20	549	41.0	716	4	US-09-171-945-125
21	487.5	36.4	419	1	US-07-649-591B-5
22	487.5	36.4	419	1	US-08-277-540-5
23	487.5	36.4	419	1	US-08-430-787A-5
24	483	36.1	417	1	US-07-649-591B-8
25	483	36.1	417	1	US-08-277-540-8
26	483	36.1	417	1	US-08-430-787A-8
27	465.5	34.8	419	4	US-08-640-906-2

28	465.5	34.8	419	4	US-09-395-936-2	Sequence 2, Appl
29	460.5	34.4	419	4	US-08-640-906-17	Sequence 17, Appl
30	460.5	34.4	419	4	US-09-395-936-17	Sequence 17, Appl
31	459	34.3	417	4	US-08-640-906-4	Sequence 4, Appl
32	459	34.3	417	4	US-09-395-936-4	Sequence 4, Appl
33	458	34.2	417	4	US-08-640-906-18	Sequence 18, Appl
34	458	34.2	417	4	US-09-395-936-18	Sequence 18, Appl
35	96.5	7.2	734	4	US-08-706-216-2	Sequence 2, Appl
36	95	7.1	706	1	US-08-484-105-16	Sequence 16, Appl
37	95	7.1	706	1	US-08-484-106-16	Sequence 16, Appl
38	90.5	6.8	434	1	US-08-111-939-13	Sequence 13, Appl
39	90.5	6.8	434	1	US-09-233-989-7	Sequence 7, Appl
40	90.5	6.8	435	1	US-08-111-939-14	Sequence 14, Appl
41	89.5	6.7	476	4	US-09-233-989-3	Sequence 3, Appl
42	89.5	6.7	476	4	US-09-233-989-6	Sequence 6, Appl
43	88.5	6.6	435	1	US-08-111-939-16	Sequence 16, Appl
44	88.5	6.6	435	1	US-08-452-262-2	Sequence 2, Appl
45	88.5	6.6	435	1	US-08-734-550-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: NO. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELETYPE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 5.5e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 ASASYEYQYHSLNEIYSWIEFTTERHPDMLTKIHGSSFEKXPYLVKXSGEQTKNNAI 60
DB 115 ASASYEYQYHSLNEIYSWIEFTTERHPDMLTKIHGSSFEKXPYLVKXSGEQTKNNAI 174
```

QY 61 WIDCIIHAREWISPAFCIMFIGH----- 83
Db 175 WIDCIIHAREWISPAFCIMFIGHITQFYGIQYTNLLRLVDYFVMPVNVNDGYDSMK 234
QY 84 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 143
Db 235 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 294
QY 144 FLRRNIQIKAYISMHSYQHIVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 295 FLRRNIQIKAYISMHSYQHIVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354
QY 204 HGHSEETLYLAPGGDDMIYDLGIKYSFT 232
Db 355 HGHSEETLYLAPGGDDMIYDLGIKYSFT 383

RESULT 2
US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-277-540-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 5.5e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 ASASYEQYHSLNETYSWIEFITERHPDMLTKIHIGSFEEKYPLVYLKSGKEQTAKNAI 60
|||||

Db 115 ASASYEQYHSLNETYSWIEFITERHPDMLTKIHIGSFEEKYPLVYLKSGKEQTAKNAI 174
QY 61 WIDCIIHAREWISPAFCIMFIGH----- 83
Db 175 WIDCIIHAREWISPAFCIMFIGHITQFYGIQYTNLLRLVDYFVMPVNVNDGYDSMK 234
QY 84 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 143
Db 235 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 294
QY 144 FLRRNIQIKAYISMHSYQHIVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 295 FLRRNIQIKAYISMHSYQHIVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354
QY 204 HGHSEETLYLAPGGDDMIYDLGIKYSFT 232
Db 355 HGHSEETLYLAPGGDDMIYDLGIKYSFT 383

RESULT 3
US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-430-787A-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 5.5e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

Qy	1	ASASYEDQYHSLNLTYSWIEETITRRHPDMLTKIHIGSSFEPYLYLVKSGKEBTAAAI	60
Db	115	ASASYEDQYHSLNLTYSWIEETITRRHPDMLTKIHIGSSFEPYLYLVKSGKEBTAAAI	174
Qy	61	WIDGCIHAREMISPAFLMTIGH-----	83
Db	175	WIDGCIHAREMISPAFLMTIGHITTOFYGLIGQYTNLRLVDFYMPVNVNDGYSMK	234
Qy	84	NRMRKNKSPFANNHCIGTDLRNRPASKHMCDEGASSSCEYTCGLYPSEPPVKAVAS	143
Db	235	NRMRKNKSPFANNHCIGTDLRNRPASKHMCDEGASSSCEYTCGLYPSEPPVKAVAS	294
Qy	144	FLRRNINQIKAYISMHSYSHIVPEYSTRSKSDHEELSLVASEAVRALEKTSKNTRYT	203
Db	295	FLRRNINQIKAYISMHSYSHIVPEYSTRSKSDHEELSLVASEAVRALEKTSKNTRYT	354
Qy	204	HGHGSETLYLAPGGDDWITDGLGKISFT	232
Db	355	HGHGSETLYLAPGGDDWITDGLGKISFT	383

```

1      RESULT      4
2      US-08-869-057-2
3      : Sequence 2, Application US/08869057
4      : Patent No. 5985562
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Morser, Michael J
8      : APPLICANT: Nagashima, Mariko
9      : TITLE OF INVENTION: Method of Detecting Thrombotic Disease
10     : TITLE OF INVENTION: Risk
11     : NUMBER OF SEQUENCES: 6
12     :
13     : CORRESPONDENCE ADDRESS:
14     : ADDRESSEE: Berlex Biosciences Legal Department
15     : STREET: 15049 San Pablo Avenue
16     : CITY: Richmond
17     : STATE: California
18     : COUNTRY: USA
19     : ZIP: 94804-0099
20     :
21     : COMPUTER READABLE FORM:
22     : MEDIUM TYPE: Floppy disk
23     : COMPUTER: IBM PC compatible
24     : OPERATING SYSTEM: PC-DOS/MS-DOS
25     : SOFTWARE: Patentln Release #1.0, Version #1.30
26     :
27     : CURRENT APPLICATION DATA:
28     : APPLICATION NUMBER: US/08/869,057
29     : FILING DATE: 03-JUN-1997
30     :
31     : CLASSIFICATION: 435
32     :
33     : ATTORNEY/AGENT INFORMATION:
34     : NAME: Wasnltien, Wendy L
35     : REGISTRATION NUMBER: 36,301
36     : REFERENCE/DOCKET NUMBER: 51509AUSM1
37     :
38     : TELECOMMUNICATION INFORMATION:
39     : TELEPHONE: 510-262-5411
40     : TELEFAX: 510-262-7095
41     :
42     : INFORMATION FOR SEQ ID NO: 2:
43     :
44     : SEQUENCE CHARACTERISTICS:
45     : LENGTH: 423 amino acids
46     : TYPE: amino acid
47     : STRANDEDNESS:
48     : TOPOLOGY: linear
49     :
50     : MOLECULE TYPE: protein
51     : ORIGINAL SOURCE:
52     : TISSUE TYPE: Plasma
53     :
54     : FEATURE:
55     : NAME/KEY: Peptide
56     : LOCATION: 23..401
57     :
58     : US-08-869-057-2

```

QY	1	ASASYEDYHSLNLTYSMIEITERRHDMMLKIHGSEFEKPYLYVKVSGKEOTANAI	60
Db	115	ASASYEDYHSLNLTYSMIEITERRHDMMLKIHGSEFEKPYLYVKVSGKEOTANAI	174
QY	61	WIDCGIHAREMISPAFLCMFIGH-----	83
Db	175	WIDCGIHAREMISPAFLCMFIGHITQFYGIGIYQYTNLLRLVDYFVMPVYVNDGYDSMK	234
QY	84	NRMWRKNSFFANNHCIGTDNRNPAASKHMCBEASSSGSEFYCGIYPSEPEPVKVAS	143
Db	235	NRMWRKNSFFANNHCIGTDNRNPAASKHMCBEASSSGSEFYCGIYPSEPEPVKVAS	294
QY	144	FLRRNINQIKAYISMHSYQHIIVPYSYTRSKSDHELSLVASEAVRALEKTSKNTRYT	203
Db	295	FLRRNINQIKAYISMHSYQHIIVPYSYTRSKSDHELSLVASEAVRALEKTSKNTRYT	354
QY	204	HGHGSEFTLYLAPGGGDDMIYDLGIKTSYFT	232
Db	355	HGHGSEFTLYLAPGGGDDMIYDLGIKTSYFT	383

```

: RESULT 5
: US-08-782-760-6
: Sequence 6, Application US/08782760
: Patent No. 5948668
: GENERAL INFORMATION:
: APPLICANT: Hartman, Jacob
: APPLICANT: Fulga, Netta
: APPLICANT: Mendelovitch, Simona
: APPLICANT: Gorecki, Marian
: TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
: TITLE OF INVENTION: CARBOXYPEPTIDASE B
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/782,760
: FILING DATE: 13-JAN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/378,233
: FILING DATE: 25-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 0336/43847
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 307 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-782-760-6

```

```

QY 1 ASASYEQVHSLNEIYSWIFTERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAI 60
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 ASHSTKRYNNMETIEMAOVATNDPDLVTOGIVGTTFEGRMYVLKI-GKTRPKPAI 59
QY 61 WIDCGIHAREMISPAFCMLFI-----GH-----83
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 FIDCGHAREMISPAFCOMFVRAVATYNOEIMKOLLDELDFYVLPVYNIDGYVTWTK 119
QY 84 NRMWRKRSFYANNHCIGDILNRNFASKHWCCEGASSSCSETYCGLYPESEPEYKAVAS 143
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 DRMRKTRSTMGSSCLGDPNRRNF-NAGMCEVGASRSPCSETYCGPAPASEKETKALAD 178
QY 144 FLRRNIQIKAYISMSHYSOHIVFPYSYTRSKSKDHEELSLVASEVRAIEKTSKNTRYT 203
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 FIRNNLSITKALITLHISQOMLIPYSYDKLPENYEELNALVGAKEEL-ATLHGTYT 237
QY 204 HGHGSETLYLAPGGDDMIYDGLIKYSFT 232
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 YGPGATTIYPAGGSDWMSYDGIKYSFT 266

```

RESULT 6

```

PCT-US96-00995-6
; Sequence 6, Application PC/TUS9600995
; GENERAL INFORMATION:
; APPLICANT: Bio-Technology General Corp.
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00995
; FILING DATE: 25-JAN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 308 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-00995-6

```

Query Match 45.6%; Score 610; DB 5; Length 307;
 Best Local Similarity 45.0%; Pred. No. 1,8e-58;
 Matches 121; Conservative 43; Mismatches 65; Indels 40; Gaps 5;

```

QY 1 ASASYEQVHSLNEIYSWIFTERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAI 60
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 ASHSTKRYNNMETIEMAOVATNDPDLVTOGIVGTTFEGRMYVLKI-GKTRPKPAI 59
QY 61 WIDCGIHAREMISPAFCMLFI-----GH-----83
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 FIDCGHAREMISPAFCOMFVRAVATYNOEIMKOLLDELDFYVLPVYNIDGYVTWTK 119

```

```

QY 84 NRMWRKRSFYANNHCIGDILNRNFASKHWCCEGASSSCSETYCGLYPESEPEYKAVAS 143
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 DRMRKTRSTMGSSCLGDPNRRNF-NAGMCEVGASRSPCSETYCGPAPASEKETKALAD 178
QY 144 FLRRNIQIKAYISMSHYSOHIVFPYSYTRSKSKDHEELSLVASEVRAIEKTSKNTRYT 203
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 FIRNNLSITKALITLHISQOMLIPYSYDKLPENYEELNALVGAKEEL-ATLHGTYT 237
QY 204 HGHGSETLYLAPGGDDMIYDGLIKYSFT 232
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 YGPGATTIYPAGGSDWMSYDGIKYSFT 266

```

RESULT 7

```

US-08-696-139-4
; Sequence 4, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hersberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-696-139-4

```

Query Match 43.5%; Score 582; DB 1; Length 306;
 Best Local Similarity 44.3%; Pred. No. 2.1e-55;
 Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

```

QY 6 YEQVHSLNEIYSWIFTERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWIWIDCG 65
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 YEKYNMETIEMATKOVSENPDILSRALGTFLGNMIVILKV-GKPGPKPAIFMDCG 64
QY 66 IHAREMISPAFCMLFI-----GH-----NRMWR 88
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 FHAREMISPAFCOMFVRAVATLYGESHMTEFLNKLDLDFYVLPVYNIDGYVTWTKNRMWR 124

```



```

QY 61 WIDCGIHAREMISPAFCIMFI-----GH----- 83
DB 168 FMDGCGHAHEMISPAFCIMFVREAVRTYGREIOVTELDKLDYVPLVINDIGYITWTX 227
QY 84 NRMNRNRSPYANNHICIGDILNRFASKHMCCEGASSCSSEYCYGLYPESEPEVAVAS 143
DB 228 SRWRRTSRTHSGSSCIGDIPNRNF-DAGWCELGASRNCDEYTCGPAASEKETKALAD 286
QY 144 ELRRNINQIKAYISMSYSQHIIVFPYSYRSKSKDHEELSLVASEAVRAIEKTSKNTXT 203
DB 287 FIRNKLSSIKAVLTIHSYSQMMIYPYSYKLGENNALAKAVKTEL-ASLHGTXT 345
QY 204 HCHGSEFLVAPGGDDWYDLGIKYSF 232
DB 346 YGPGATTIYPNAGSDDMAYDGIIRISFT 374

RESULT 10
US-07-649-591B-7
; Sequence 7, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haesk, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
US-07-649-591B-7

Query Match 41.9%; Score 561; DB 1; Length 417;
Best Local Similarity 43.7%; Pred. No. 6,5e-53;
Matches 115; Conservative 37; Mismatches 71; Indels 40; Gaps 5;
QY 6 YEOYHSINETYNIETFERHPMLTKIHIGSFEKYPPIYLVKSGKEQTAKNAIWDG 65
DB 116 YAKYNDMDKTVSWEKLEKHPVSRIRIKISTVEDNPLVTKI-GKKGDERRAIFMDG 174
QY 66 IHAREMISPAFCIMFI-----GH-----RMR 88
DB 175 IHAREMISPAFCIMFYQATKSGKKNKIMTKLLDRANFYVLPVFNVDGYTWSVTDQMR 234
QY 89 KNSFYANNHICIGDILNRFASKHMCCEGASSCSSEYCYGLYPESEPEVAVASFLRN 148

```

```

DB 235 KNSRNRNOSTCIGDILNRF-DVSWDSSPNINKPCLNRYRGAPSEKETRAVNFISH 293
QY 149 INQIAYISMSYSQHIIVFPYSYRSKSKDHEELSLVASEAVRAIEKTSKNTRTYHGS 208
DB 294 LNSIKAYITFHSYSQMLIPYGTFLPNNODLKVARIATDAL-SRYEYRRTYGPJA 352
QY 209 EFLVAPGGDDWYDLGIKYSF 231
DB 353 STYIKTSSGLDWMYDLGIKRTE 375

```

```

RESULT 11
US-08-277-540-7
; Sequence 7, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Haesk, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689DICI1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-277-540-7

```

```

Query Match 41.9%; Score 561; DB 1; Length 417;
Best Local Similarity 43.7%; Pred. No. 6,5e-53;
Matches 115; Conservative 37; Mismatches 71; Indels 40; Gaps 5;
QY 6 YEOYHSINETYNIETFERHPMLTKIHIGSFEKYPPIYLVKSGKEQTAKNAIWDG 65
DB 116 YAKYNDMDKTVSWEKLEKHPVSRIRIKISTVEDNPLVTKI-GKKGDERRAIFMDG 174
QY 66 IHAREMISPAFCIMFI-----GH-----RMR 88
DB 175 IHAREMISPAFCIMFYQATKSGKKNKIMTKLLDRANFYVLPVFNVDGYTWSVTDQMR 234

```


Db 223 SCIGVRNRNF-NAGWCEVGASRSPCSETTTCGPAPESKEKTKALADIFRNNLSTIKATLT 281
QY 158 MHSYSQIIVPEPYSTRSKSDHELSLVASAVAIKTSKNTRYTHGSGETLYLAPGG 217
Db 282 IHSYSQMLPYSDYKLPENYEELNALVGAAREL-ATLHGTYTEPGATITYPAAAG 340
QY 218 GDDWITYDLGKYSFT 232
Db 341 SDDWSYDGIKYSFT 355

RESULT 14

US-08-277-540-4
; Sequence 4, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA: 07/959944
; FILING DATE: 14-OCT-1992
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-277-540-4

Query Match 41.4%; Score 554; DB 1; Length 396;
Best Local Similarity 44.3%; Pred. No. 3.5e-52;
Matches 113; Conservative 39; Mismatches 63; Indels 40; Gaps 5;

QY 15 IYSWIEFTEHRHPDMLTKIHGSSFEKYPYLVKSGKEQTAKNAVMDGCIHAREWISP 74
Db 104 IEAMIQOAVATNPDLVQSVIGTFEGRNMYVLKI-GKTRNKPATIFIDCGFHAREWISP 162
QY 75 AFCLMFI-----GH-----NRMWRKNSFYANN 97
Db 163 AFCLMFI-----GH-----NRMWRKNSFYANN 97
QY 98 HCLGTDLNRFASKHMCCEGASSSSCSETYCGLYPESEPEYKAVASFLRNNINQIKAYIS 157

Db 223 SCIGVRNRNF-NAGWCEVGASRSPCSETTTCGPAPESKEKTKALADIFRNNLSTIKATLT 281
QY 158 MHSYSQIIVPEPYSTRSKSDHELSLVASAVAIKTSKNTRYTHGSGETLYLAPGG 217
Db 282 IHSYSQMLPYSDYKLPENYEELNALVGAAREL-ATLHGTYTEPGATITYPAAAG 340
QY 218 GDDWITYDLGKYSFT 232
Db 341 SDDWSYDGIKYSFT 355

RESULT 15

US-08-430-787A-4
; Sequence 4, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-430-787A-4

Query Match 41.4%; Score 554; DB 1; Length 396;
Best Local Similarity 44.3%; Pred. No. 3.5e-52;
Matches 113; Conservative 39; Mismatches 63; Indels 40; Gaps 5;

QY 15 IYSWIEFTEHRHPDMLTKIHGSSFEKYPYLVKSGKEQTAKNAVMDGCIHAREWISP 74
Db 104 IEAMIQOAVATNPDLVQSVIGTFEGRNMYVLKI-GKTRNKPATIFIDCGFHAREWISP 162
QY 75 AFCLMFI-----GH-----NRMWRKNSFYANN 97

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:44:08 ; Search time 55.22 Seconds

(without alignments)
1127.819 Million cell updates/sec

Title: us-09-980-881-2

Perfect score: 1911

Sequence: 1 MKLCSLAVLPVIFVFCQHV.....IKYSFTSNPYEKLPLSLK 360

Scoring table: BLOSUM62

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEA:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1884	98.6	360	4	Q9P2Y6
2	1799.5	94.2	423	4	Q15114
3	1794.5	93.9	423	4	Q961Y4
4	1528	80.0	422	11	Q9JH6
5	1522	79.6	422	11	Q9QZFO
6	1499	78.4	422	11	Q9QZFO
7	1029	53.8	198	4	Q9NRI8
8	700.5	36.7	416	13	Q9PUF2
9	646	33.8	416	6	Q9XSP3
10	641	33.5	417	4	Q96B08
11	624.5	32.7	417	4	Q96B08
12	621	32.5	412	11	P97597
13	531	27.8	279	11	Q9CVD1
14	516.5	27.0	419	6	Q9TV85
15	511	26.7	417	4	Q96GN3
16	511	26.7	417	4	Q96A12

17	489.5	25.6	419	4	Q9B567	Q9B567	homo sapien
18	471.5	24.7	247	4	Q96K29	Q96K29	homo sapien
19	460	24.1	424	5	Q9V187	Q9V187	homo sapien
20	459	24.0	1192	5	Q9W475	Q9W475	homo sapien
21	451.5	23.6	422	5	Q9V122	Q9V122	homo sapien
22	436	22.8	584	5	Q9T2H1	Q9T2H1	homo sapien
23	418	21.9	440	5	Q9W478	Q9W478	homo sapien
24	407.5	21.3	419	5	Q9V121	Q9V121	homo sapien
25	402	21.0	424	5	Q9V566	Q9V566	homo sapien
26	400.5	21.0	400	5	Q9V186	Q9V186	homo sapien
27	399.5	20.9	427	5	Q9V9K2	Q9V9K2	homo sapien
28	397	20.8	437	5	Q9V9K2	Q9V9K2	homo sapien
29	395.5	20.7	540	5	Q9T2H2	Q9T2H2	homo sapien
30	391	20.5	453	5	Q9V342	Q9V342	homo sapien
31	385	20.1	423	5	Q9V132	Q9V132	homo sapien
32	382.5	20.0	581	5	Q9V338	Q9V338	homo sapien
33	370	19.4	545	5	Q9V356	Q9V356	homo sapien
34	367	19.2	1430	5	Q9VX86	Q9VX86	homo sapien
35	366	19.2	467	5	Q9VCM8	Q9VCM8	homo sapien
36	362.5	19.0	455	5	Q9VCM8	Q9VCM8	homo sapien
37	358	18.7	666	5	Q9VX75	Q9VX75	homo sapien
38	355	18.6	415	5	Q9V564	Q9V564	homo sapien
39	349.5	18.3	400	5	Q9V567	Q9V567	homo sapien
40	344.5	18.0	354	5	Q9V568	Q9V568	homo sapien
41	344.5	18.0	418	5	Q9V1J8	Q9V1J8	homo sapien
42	337	17.6	312	5	Q9VR23	Q9VR23	homo sapien
43	333	17.4	315	4	Q9NR19	Q9NR19	homo sapien
44	324.5	17.0	161	13	Q90W12	Q90W12	homo sapien
45	320	16.7	418	3	Q14418	Q14418	metathizium

ALIGNMENTS

RESULT 1
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCF-2000 (TREMREL. 15, Created)
DT 01-OCF-2000 (TREMREL. 15, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE CARBOXYPEPTIDASE B-LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; AB011869; BAA90475.1; -.
DR HSSP; P00730; ZCTC.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypep.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypep; 2.
DR PRINTS; PR00765; CRBOXPPTASEA.
DR PROSITE; PS00765; B670B2F7437C1CB CRC64;
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

NO.

Query Match 98.6%; Score 1884; DB 4; Length 360;
Best Local Similarity 98.3%; Pred. NO. 2.2e-162;
Matches 354; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 MKLCSLAVLPVIVLFCEDHVFAGFQSGVLAALPRTSRQVLOLNTTTEIYVLMQPTAD 60
DB 1 MKLCSLAVLPVIVLFCEDHVFAGFQSGVLAALPRTSRQVLOLNTTTEIYVLMQPTAD 60
QY 61 LIYKKQVHFFVNASDVNDVNAKHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
DB 61 LIYKKQVHFFVNASDVNDVNAKHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
QY 121 EGYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVSCKGKOTAKNAIWDGCI 180
DB 121 EGYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVSCKGKOTAKNAIWDGCI 180
QY 181 HAREWISPAFCLEFIGNHRRMRKNSRFFYANNHCIGTDLNSNFVSKHWCCEGASSSCSET 240
DB 181 HAREWISPAFCLEFIGNHRRMRKNSRFFYANNHCIGTDLNSNFVSKHWCCEGASSSCSET 240
QY 241 YCGLYPESEPEVKAASFLLRNINQIKAYISMHSYSHIYFPYSYTRSKSDHEELSLVA 300
DB 241 YCGLYPESEPEVKAASFLLRNINQIKAYISMHSYSHIYFPYSYTRSKSDHEELSLVA 300
QY 301 SEAVRAIDKTSKNTRYTHGHSESTLYLAPGCGDDWYIDLGIKYSFNSPVEKILPLSLK 360
DB 301 SEAVRAIDKTSKNTRYTHGHSESTLYLAPGCGDDWYIDLGIKYSFNSPVEKILPLSLK 360
RESULT 2
Q05114 PRELIMINARY: PRT: 423 AA.
AC 015114:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PCPB PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Mallory B.E., Tsai S.P., Henzel W., Dryden D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21835-21838(1991).
DR EMBL; M75106; AAA60042.1; -.
DR HSSP; P00730; 1PYT.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Proper_M14.
DR InterPro; IPR000834; Zn_carboxypep.
DR Pfam; PF02246; Proper_M14; 1.
DR PRINTS; PR00765; ZN_CARBOXYPEP.1.
DR CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6EE79F4 CRC64;

```

Full length

```

DB 121 EGYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVSCKGKOTAKNAIWDGCI 180
QY 181 HAREWISPAFCLEFIGNHRRMRKNSRFFYANNHCIGTDLNSNFVSKHWCCEGASSSCSET 240
DB 181 HAREWISPAFCLEFIGNHRRMRKNSRFFYANNHCIGTDLNSNFVSKHWCCEGASSSCSET 240
QY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESPEVKAASFLLRNI 263
DB 241 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESPEVKAASFLLRNI 300
QY 264 NQIKAYISMHSYSHIYFPYSYTRSKSDHEELSLVA SEAVRAIDKTSKNTRYTHGHSE 323
DB 301 NQIKAYISMHSYSHIYFPYSYTRSKSDHEELSLVA SEAVRAIDKTSKNTRYTHGHSE 360
QY 324 TLYLAPGCGDDWYIDLGIKYSFT 346
DB 361 TLYLAPGCGDDWYIDLGIKYSFT 383
RESULT 3
Q061Y4 PRELIMINARY: PRT: 423 AA.
AC 0961Y4:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE CARBOXYPEPTIDASE B2 (PLASMA).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELETAL MUSCLE;
RA Strauberg R.;
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC007057; AA07057.1; -.
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6EE7CF5 CRC64;

```

```

Query Match 93.9%; Score 1794.5; DB 4; Length 423;
Best Local Similarity 89.3%; Pred. No. 3,6e-154;
Matches 342; Conservative 1; Mismatches 3; Indels 37; Gaps 1;
QY 1 MKLCSLAVLPVIVLFCEDHVFAGFQSGVLAALPRTSRQVLOLNTTTEIYVLMQPTAD 60
DB 1 MKLCSLAVLPVIVLFCEDHVFAGFQSGVLAALPRTSRQVLOLNTTTEIYVLMQPTAD 60
QY 61 LIYKKQVHFFVNASDVNDVNAKHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
DB 61 LIYKKQVHFFVNASDVNDVNAKHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
QY 121 EGYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVSCKGKOTAKNAIWDGCI 180
DB 121 EGYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVSCKGKOTAKNAIWDGCI 180
QY 181 HAREWISPAFCLEFIGNHRRMRKNSRFFYANNHCIGTDLNSNFVSKHWCCEGASSSCSET 240
DB 181 HAREWISPAFCLEFIGNHRRMRKNSRFFYANNHCIGTDLNSNFVSKHWCCEGASSSCSET 240
QY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESPEVKAASFLLRNI 263
DB 241 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESPEVKAASFLLRNI 300
QY 264 NQIKAYISMHSYSHIYFPYSYTRSKSDHEELSLVA SEAVRAIDKTSKNTRYTHGHSE 323
DB 301 NQIKAYISMHSYSHIYFPYSYTRSKSDHEELSLVA SEAVRAIDKTSKNTRYTHGHSE 360
QY 324 TLYLAPGCGDDWYIDLGIKYSFT 346
DB 361 TLYLAPGCGDDWYIDLGIKYSFT 383

```



```

RESULT 4
ID 09JHH6 PRELIMINARY; PRT; 422 AA.
AC 09JHH6;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE R (THROMBIN-ACTIVATABLE FIBRINOLYSIS INHIBITOR)
GN (111003P04RIK PROTEIN).
DE CPB2 OR TAFI OR 111003P04RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20341711; PubMed=10878383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Odate K., Okada N.,
RA Campbell W., Okada H.;
RA "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not.";
RT J. Immunol. 165:1053-1058(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20201996; PubMed=10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor.";
RT Thromb. Haemost. 83:297-303(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochina H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Sakai L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilmink L.,
RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RL Nature 409:685-690(2001).
DR EMBL; AB021968; BAB03402.1; -
DR EMBL; AF164524; AAF62385.1; -
DR EMBL; AK004045; BAB23141.1; -
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

```

Query Match 80.0%; Score 1528; DB 11; Length 422;
 Best Local Similarity 76.2%; Pred. No. 5e-130;
 Matches 292; Conservative 24; Mismatches 29; Indels 38; Gaps 2;

QY 1 MKLCSLAVIVPIYLFCEQHAFQSGOVLALPRTSRQVLOLNTTTEIVLMQPVYAD 60

```

DB 1 MKLHGLILVAILLY-EGHGFQSGOVLALPRTSRQVLOLNTTTEIVLMQPVYAE 59
QY 61 LIYKKQVHFVNASDVNDVNAKHLNVSGIPCSVLADVEDLIQQOISNDVSPRASASY 120
DB 60 FIEKKKEHFVNASDVSDVNAKHLNVSKRIPPNVLMNVEDLIEQOTFNDVSPRASASY 119
QY 121 EGYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKPYLVYLVKSGKEOTANAIWIDCGI 180
DB 120 EGYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKPYLVYLVKSGKEORIKNAIWDGCI 179
QY 181 HAREWISPAFCWFIH-----NRMRK 203
DB 180 HAREWISPAFCWFIHGYVTOFHGREKNTYRLRHVDFYIMFVNVNDGYDTWKKNRMWRK 239
QY 204 NRSFYANNHCIGTDLNENFNSKHWCEGASSSCSEYTCGIPSEPEVKAVALRNI 263
DB 240 NRSFANNRCVGTDLNRFNSKHWCEGASSSCSEYTCGIPSEPEVKAVALRNI 299
QY 264 NQIKAYISMHSQHIYFYPYSYTRSKSDHEELSLVSEAVRAIDKTSKNTRYTHGSE 323
DB 300 DHIKAYISMHSYQQLIFPYSYTRSKSDHEELSLVSEAVRAIESINKNTRYTHGSGSE 359
QY 324 TLYLAPGGDDMYIDLGIKYSFT 346
DB 360 SLYLAPGGSDMYIDLGIKYSFT 382

```

Query Match 79.6%; Score 1522; DB 11; Length 422;
 Best Local Similarity 76.0%; Pred. No. 1.7e-129;
 Matches 291; Conservative 24; Mismatches 30; Indels 38; Gaps 2;

QY 1 MKLCSLAVIVPIYLFCEQHAFQSGOVLALPRTSRQVLOLNTTTEIVLMQPVYAD 60
 DB 1 MKLHGLILVAILLY-EGHGFQSGOVLALPRTSRQVLOLNTTTEIVLMQPVYAE 59
 QY 61 LIYKKQVHFVNASDVNDVNAKHLNVSGIPCSVLADVEDLIQQOISNDVSPRASASY 120
 DB 60 FIEKKKEHFVNASDVSDVNAKHLNVSKRIPPNVLMNVEDLIEQOTFNDVSPRASASY 119
 QY 121 EGYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKPYLVYLVKSGKEOTANAIWIDCGI 180

```

Db 120 EQYSLNEIYSWIEVFTEQHPDMLOKITYIGSSFEKPYLYLKVGSKQRIKNAIWDGCI 179
QY 181 HAREWISPAFLMEIGH-----NRMRK 203
Db 180 HAREWISPAFLMEIGHVYVTFEGHGENLYTRLRHVDYIIMPVNVVDYDTTWKKNMRK 239
QY 204 NRSFYANNHCIGDNLNENFSKHKCEGASSSCSEFYGLYPESEPEKAVASFLRNTI 263
Db 240 NRSAHKNNRCVGTDLNBNFASKHCEKASSSCSEFYCGLYPESEPEKAVADFLRNTI 299
QY 264 NQIKAYISMHSYQIILFPYSYRNSKSDHELSLVAESAIVRAIGSFNNTRYTHGSGSE 323
Db 300 DHIKAYISMHSYQIILFPYSYRNSKSDHELSLVAESAIVRAIGSFNNTRYTHGSGSE 359
QY 324 TLYLAPGGDDMIYDGIKYSFT 346
Db 360 SLYLAPGGSDMIYDGIKYSFT 382

```

RESULT 6

```

Q9EOV9 PRELIMINARY; PRT; 422 AA.
AC Q9EOV9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PRE-PROCARBOXYPEPTIDASE R.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20471387; PubMed=11021404;
RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA Okada N., Okada H.;
RT "Molecular cloning and partial characterization of rat
RT procarboxypeptidase R and carboxypeptidase N.";
RL Microbiol. Immunol. 44:719-728(2000).
DR EMBL: AB042598; BAB18617.1; -.
DR HSRP: P00730; ZC7C.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 1.
DR PRINTS: PR00765; CRBOXYPTRASEA.
KW Carboxypeptidase.
SQ SEQUENCE 422 AA; 48826 MW; FFFD32A51A9366C8 CRC64;

```

Query Match 78.4%; Score 1499; DB 11; Length 422;
 Best Local Similarity 74.7%; Pred. No. 2.1e-127;
 Matches 286; Conservative 26; Mismatches 33; Indels 38; Gaps 2;

```

QY 1 MKLCSLAVLPVILFCEQHVFAFSGOVLALPRTSQOVLYOMLTYYEIVLMQPYTAD 60
Db 1 MKLYGLVLAIVILY-EKHGLAFSGHVALPRTSRQVLYOMLTYYEIVLMQPYTAE 59
QY 61 LIVKKQVHFVNASDVDNKKAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
Db 60 FIEKKKEVHFVNASDVNKAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 119
QY 121 EQYSLNEIYSWIEFTEHRHPDMLTIKIHGSSFEKPYLYLKVGSKQRTAKNAIWDGCI 180
Db 120 EQYSLNEIYSWIEFTEHRHPDMLOKITYIGSSFEKPYLYLKVGSKQRTAKNAIWDGCI 179
QY 181 HAREWISPAFLMEIGH-----NRMRK 203
Db 180 HAREWISPAFLMEIGHVYVTFEGHGENLYTRLRHVDYIIMPVNVVDYDTTWKKNMRK 239
QY 204 NRSFYANNHCIGDNLNENFSKHKCEGASSSCSEFYCGLYPESEPEKAVASFLRNTI 263

```

```

Db 240 NRSVHMNNRCVGTDLNBNFASKHCEKASSSCSEFYCGLYPESEPEKAVADFLRNTI 299
QY 264 NQIKAYISMHSYQIILFPYSYRNSKSDHELSLVAESAIVRAIDKTSKNTRYTHGSGSE 323
Db 300 NHTKAYISMHSYQIILFPYSYRNSKSDHELSLVAESAIVRAISINKNTRYTHGSGSE 359
QY 324 TLYLAPGGDDMIYDGIKYSFT 346
Db 360 SLYLAPGGSDMIYDGIKYSFT 382

```

RESULT 7

```

Q9NTI8 PRELIMINARY; PRT; 198 AA.
AC Q9NTI8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BA139H14.2 (CARBOXYPEPTIDASE B2 (PLASMA)) (FRAGMENT).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tromans A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL137141; CAB92622.1; -.
DR HSRP: P00730; 1CPX.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 1.
DR PRINTS: PR00765; CRBOXYPTRASEA.
KW Carboxypeptidase.
FT NOK_TER 198
SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;

```

Query Match 53.8%; Score 1029; DB 4; Length 198;
 Best Local Similarity 99.5%; Pred. No. 2.7e-85;
 Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MKLCSLAVLPVILFCEQHVFAFSGOVLALPRTSQOVLYOMLTYYEIVLMQPYTAD 60
Db 1 MKLCSLAVLPVILFCEQHVFAFSGOVLALPRTSQOVLYOMLTYYEIVLMQPYTAD 60
QY 61 LIVKKQVHFVNASDVDNKKAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
Db 61 LIVKKQVHFVNASDVDNKKAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASY 120
QY 121 EQYSLNEIYSWIEFTEHRHPDMLTIKIHGSSFEKPYLYLKVGSKQRTAKNAIWDGCI 180
Db 121 EQYSLNEIYSWIEFTEHRHPDMLOKITYIGSSFEKPYLYLKVGSKQRTAKNAIWDGCI 180
QY 181 HAREWISPAFLMEIGH 197
Db 181 HAREWISPAFLMEIGH 197
QY 181 HAREWISPAFLMEIGH-----NRMRK 203
Db 181 HAREWISPAFLMEIGH-----NRMRK 203
QY 204 NRSFYANNHCIGDNLNENFSKHKCEGASSSCSEFYCGLYPESEPEKAVASFLRNTI 263
Db 204 NRSFYANNHCIGDNLNENFSKHKCEGASSSCSEFYCGLYPESEPEKAVASFLRNTI 263

```

```

OX  NCBI_TaxID=8724;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT  Screening of Bothrops jararaca pancreas cDNA library.
RL  Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AF190274; AAF01344.1; -.
DR  HSSP: P09955; INSA.
DR  MEROPS: M14.003; -.
DR  InterPro: IPR003146; Propep.M14.
DR  InterPro: IPR000834; Zn_carboxypept.
DR  Pfam: PF02244; Propep.M14; 1.
DR  Pfam: PF00246; Zn_carboxypept.1.
DR  PRINTS: PR00765; CARBOXYPTASEA.
DR  PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR  PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
SQ  SEQUENCE 416 AA; 47723 MW; 4F999854DD72B7A7 CRC64;

Query Match          36.7%; Score 700.5; DB 13; Length 416;
Best Local Similarity 38.5%; Pred. No. 4,5e-55;
Matches 142; Conservative 65; Mismatches 119; Indels 43; Gaps 5;

OY  15 FCGHVFAGSGOVLALPRTSRQOVLQNLTTTTEIVLMQPTADLIYKKQHFYVNA 74
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  14 FAFTYHREDGEKYYRTPEDEVTFLNTANIVQDFWRPDSVELYKAKMTDFERIEA 73
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  75 SDVNTYKALHNVSGIPCSVLLADVEDLIQQOISNDTVSPRASAYYQVHSLNEIYSMIE 134
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  74 DRCSEVESILQSGNLNELLIDNLQAVLDRLDNH---RTAGVNTKYSWMEKIDMTA 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  135 FTERHPDMLTKHISSEFKIPLYLAKSGKEQTAKNAIWDGCHAREMISPACFLMF 194
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  131 DIANENPSLVSRLQIGTFGRPMPLKLV-GKPGVNKKAIFIDCGFARAWISPAFCQWF 189
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  195 I-----GHNRMKRKRSFYANNHCGTD 217
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  190 VRAVAVTYGKETITMOLINKLDFYILPVLNIDGYVSMKOSRMKRKTRVNASGTCIGTD 249
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  218 LNSNFVSKHNCBEGASSSCSETTCGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQ 277
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  250 PNNNF-DAAMCVSGASNPCESETTCGSGPESEKETALADFIIRNRRIIDAYLIHSYSQ 308
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  278 HLYFPYSTRSKDHELSVASEAVRAIDKSKNTRTHGSGELLYLAPGGGDWITY 337
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  309 MLYFPYSTYDILTSNNKLNLSIAKEAIRL-KVLFGETYTGCAATYPAAGSGDDWY 367
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  338 DLGIXYSFT 346
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  368 DGKITYFT 376
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
O9XSP3 PRELIMINARY; PRT; 416 AA.
AC O9XSP3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE PROCARBOXYPEPTIDASE B PRECURSOR (EC 3.4.17.2).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA MEDLINE=99321926; PubMed=10391940;
RA Ventura S., Villegas V., Steiner J., Larson J., Vendrell J.,
RA Hersberger C., Ayiles F.;
RT Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL J. Biol. Chem. 274:19925-19933(1999).

```

```

DR  EMBL: AJ133775; CAB46991.1; -.
DR  HSSP: P09955; IPBA.
DR  MEROPS: M14.003; -.
DR  InterPro: IPR003146; Propep.M14.
DR  InterPro: IPR000834; Zn_carboxypept.
DR  Pfam: PF02244; Propep.M14; 1.
DR  Pfam: PF00246; Zn_carboxypept.1.
DR  PRINTS: PR00765; CARBOXYPTASEA.
DR  PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR  PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Signal; Carboxypeptidase; Hydrolase.
FT SIGNAL 1 15
FT CHAIN 16 416
FT SEQUENCE 416 AA; 47381 MW; 5FAVFCED7B45ABEA CRC64;

Query Match          33.8%; Score 646; DB 6; Length 416;
Best Local Similarity 36.6%; Pred. No. 3.9e-50;
Matches 140; Conservative 64; Mismatches 134; Indels 44; Gaps 7;

OY  3 LCSLAVLPVLPCEQHVFA-FQSGOVLALPRTSRQOVLQNLTTTTEIVLMQPTADL 61
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  1 MIAFLILVTYVLASAHSGHFEGEKVFRVNVEDENDISLHELASTRLQDFWKPDSVTQ 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  62 IYKKQVHFYVNASDVNKAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASASTYE 121
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  61 IKPSTVDFRYKARDILAVEDLEQNELOVEVLINNLRSVLAQFDSRV--RTGHSYE 117
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  122 QVHSLNEIYSWIEFTEHRPDLTKRHGSEFKYPLYLAKSGKEQTAKNAIWDGCH 181
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  118 KYNMETTEAWTKQVTSNPDLISRTAIGTGLNNIYLLKV-GKPGPNKPAITMDGFEH 176
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  182 AREMISPAFCIMFT-----GH-----NRMRKN 204
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  177 AREMISPAFCQWFREAVLFTGYESHMTEFLNKLDYFVLPVLNIDGYLYTFTKRMKRKT 236
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  205 RSFYANNHCITDLSNFSVSKHNCBEGASSSCSETTCGLYPESEPEYKAVASFLRNIN 264
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  237 RSTNAGTCTIGTDPNRNF-DAGMCTTGASTDPCDETCGSAASESEKETKALADFIIRNLS 295
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  265 QIKAYISMHSYSHIIVPYSTRSKSDHELSVASEAVRAIDKTSKNRTHGSGSET 324
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  296 STAYLIHISTQMLTLYSIDYKLPENNALNLAAVAKEL-ATLYGTITTYGPGATT 354
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY  325 LYLAPEGGDWITYDLGIXYSFT 346
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  355 IYPAAGSGDDWYDGIKYSFT 376
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
O96B08 PRELIMINARY; PRT; 417 AA.
AC O96B08;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HYPOHETICAL 47.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADRENAL CARCINOMA;
RA Strausberg R.;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC015338; AAH15338.1; -.
KW Hypothetical protein.
SQ SEQUENCE 417 AA; 47367 MW; EBB98827F5D5AF9 CRC64;

Query Match          33.5%; Score 641; DB 4; Length 417;
Best Local Similarity 35.3%; Pred. No. 1.1e-49;

```

Matches 135; Conservative 71; Mismatches 132; Indels 44; Gaps 7;

```

QY 3 LGSIAVLVPLVFECEQVFA-FOSGOVLAALPRTSROVYONLTYYEIVLMOPVADL 61
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MLMLVLVTVLAALSAHGHGEFEKEFRVNVEDENINIRIETASTTQIDFKPDSVQ 60
QY 62 IYKKQVHFVNASDVNDVKAHLNVSIGPISVLADVEDLIQOQISNDTVSPRASASYE 121
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 IKPHSTVDFVKAEDTVENVLKQNELQYKVLISNLRNVEAOFSR---RATGSHYE 117
QY 122 QYHSLNEIYMIERFERHDMCLKIHGSSFEKYPYLVKVGSKETANAIWIDCGIH 161
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 KYKWEITEAMTOOVATENALISRSVITTEBRATYTLKV-GKACQNPALFMDCGFH 176
QY 182 AREMISPAFLMFT-----GH-----NRMARKN 204
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 AREMISPAFCOMFYREAVRTYGREIOVTELLDKIDFYVLPVLNIDGIYTWTSRFRMKT 236
QY 205 RSPFYANNHCIGTDLNLSFVSKHMCBEGASSSCSEYTCGLYPSSEPEVKAASFLLRNIN 264
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 RSTHTGSSCIGTDPNRF-DAGWCEIGASRNPCEDEYCGPMAESEKETKALADFIKRLS 295
QY 265 QIKAVISMHSYSOHIFEPYSYTRSKSDHEELSLVASEAVRAIDKSKNRTYHGHGSET 324
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 SIKAVYLIHSYSGMMIYPIYSYAKKLGENNENALAKATYKEL-ASLHGKITTYGEGATT 354
QY 325 LYIAPGGDDMYIDLGIKYSF 346
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 IYPMAGSDMDWAYDGIKYSF 376

```

RESULT 11
096E94 PRELIMINARY; PRT: 417 AA.

AC 096E94; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE SIMILAR TO CARBOXYPEPTIDASE A3 (MAST CELL).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW, AND ACUTE MYELOCTYTIC LEUKEMIA;
RA Straussberg R.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012613; AAH12613.1; -
KW Carboxypeptidase.
SQ SEQUENCE 417 AA; 48714 MW; F781EEF23F570E98 CRC64;

Query Match 32.7%; Score 624.5; DB 4; Length 417;
Best Local Similarity 34.4%; Pred. No. 3.4e-48;
Matches 131; Conservative 76; Mismatches 127; Indels 47; Gaps 8;

```

QY 6 LAVLPVILFCEQVFA---FOSGOVLAALPRTSROVYONLTYYEIVLMOPVADLI 62
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MLMLVLVTVLAALSAHGHGEFEKEFRVNVEDENINIRIETASTTQIDFKPDSVQ 60
QY 63 VKKQVHFVNASDVNDVKAHLNVSIGPISVLADVEDLIQOQISNDTVSPRASASYE 121
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 AAKMMVDFVSKESQALQASLDQKMHYELIHDLOEIEQOFVKEDIGERHS---YA 117
QY 122 QYHSLNEIYMIERFERHDMCLKIHGSSFEKYPYLVKVGSKETANAIWIDCGIH 181
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 KYKWEITEAMTOOVATENALISRSVITTEBRATYTLKV-GKACQNPALFMDCGFH 176
QY 182 AREMISPAFLMFT-----GH-----NRMARKN 204
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 AREMISPAFCOMFYQATKTYGRKNIMTKLLDRANFYLLPVFNVDGYIWSWTKNRMWKN 236

```

```

QY 205 RSPFYANNHCIGTDLNLSFVSKHMCBEGASSSCSEYTCGLYPSSEPEVKAASFLLRNIN 264
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 RSKNNSKSCIGTDLNRF-NASNNISIPNTNDPADNVRSPASESEKETRYHNFISHLN 295
QY 265 QIKAVISMHSYSOHIFEPYSYTRSKSDHEELSLVASEAVRAIDKSKNRTYHGHGSET 324
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 EIKVYLIHFHSYQMLLPEFGYTSKLPNNHDLAKAKIGTDVL-STRYETRYIYGIEST 354
QY 325 LYIAPGGDDMYIDLGIKYSF 345
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 IYPISSSLDMDWAYDGIKHTF 375

```

RESULT 12
P97597 PRELIMINARY; PRT: 412 AA.

AC P97597; 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE MAST CELL CARBOXYPEPTIDASE A PRECURSOR (FRAGMENT).
GN R-CPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY;
RX MEDLINE=97149430; PubMed=8996238;
RA Lutzelschwab C., Rejler G., Aveskogh M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations".
RL J. Exp. Med. 185:13-29(1997).
DR EMBL: U67914; AAB48267.1; -
DR HSSP; P09955; INSA.
DR MEROPS; M14.010; -
DR InterPro; IPR003146; Propep_M14.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept_1.
DR PRINTS; PR00765; CRBOXYPEPTASE.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Signal; Carboxypeptidase.
FT NON_TER 1
FT SIGNAL 1
FT CHAIN 1
FT CHAIN 11
FT CHAIN 412
SQ SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

Query Match 32.5%; Score 621; DB 11; Length 412;
Best Local Similarity 37.2%; Pred. No. 7e-48;
Matches 134; Conservative 55; Mismatches 129; Indels 42; Gaps 6;

```

QY 23 FOSGOVLAALPRTSROVYONLTYYEIVLMOPVADLIYKKQVHFVNASDVNDVKA 82
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 FDRKEVFRVKLQDEKQASILKNLTQTELDWYDPAIDHADVNNVTQFRTYKESQTIQS 75
QY 83 HLNVSIGPISVLADVEDLIQOQISNDTVSPRASASYEYQYHSLNEIYMIERFERHDP 142
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 TLEQHKMDEYELINDLEIDKQF--DVKEEDIAGRHSYAKINDNKKIVSTEKVAVERHPE 133
QY 143 MLTKIHGSSFEKYPYLVKVGSKETAKNAIWIDCGIHAREWISPAFLMFT----- 195
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 134 MYSRIKIGSTYEDNDPIYVYK-GRKDEGRKALIFMDCGIIHAREWISPAFCOMFYQAASY 192
QY 196 GHN-----RMMARKNSFYANNHCIGTDLNLSFVSK 225
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 GKKNIMTKLLDRANFYLLPVFNVDGYIWSWTKDMRMKNSKNPNSTCIGTDLNRF-DV 251
QY 226 HMCBEGASSSCSEYTCGLYPSSEPEVKAASFLLRNINIKAVISMHSYSOHIFEPYSY 285

```

```

Db 252 SMOSSPNTNPCLSVYRGAPPESEKETKAVNTIRSHINSIKAYITPHSHSOMLIPFYGT 311
QY 286 TRSKSDHEELSLVASEAVRAIDKTSKNTYTHGSESLYLAPGGDDMIYDLGIKYSF 345
Db 312 TIKLPNHODLKVARIATDVL-SSRYEPRYIGYPIASTIKYKSSGLDMAYDGLGKHTE 370

RESULT 13
Q9CYD1 PRELIMINARY; PRT; 279 AA.
AC 09CYD1
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE 221008M23RIK PROTEIN (FRAGMENT).
GN 221008M23RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Lyons F., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilmink L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RL EMBL: AK008678; BAB25826.1; -.
DR HSSP; P09955; INSA.
DR MGI: MGI:192353; 221008M23RIK.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CARBOXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; UNKNOWN_1.
FT NON_TER 279
FT SEQUENCE 279 AA; 31970 MW; 29468AFA0B67E0AA CRC64;

Query Match 27.8%; Score 531; DB 11; Length 279;
Best Local Similarity 45.6%; Pred. No. 5.7e-40;
Matches 103; Conservative 31; Mismatches 52; Indels 40; Gaps 4;

QY 158 LVYKVSQGEQRAKNAIMIDCGIHAREWISPAFCIMFI----- 195
Db 1 MYVLKI-GDRPNKPAFFIDCGFHAREWISPAFCOMFEAVTYKQEIHKRLDELDF 59
QY 196 -----GHNRMKRNRSFYANNHCIGTDLNSNFYSKHMCEGASSSCSET 240
Db 60 YVLPVNVNDICYYTAKDKRMKRTKRTSTTAGSSCFGYDPNRNF-DAGKCEGASRSCSDT 118
QY 241 YGGLPESPEPEKAVASFLRNINQIKAYISMYSQHIATFPYSTRSKSDHEELSLVA 300
Db 119 YGPFPESEKETKALADFIKONLSIKAYITVHSYQMLPYPSYDYKLPENVEELNATV 178
QY 301 SEAVRAIDKTSKNTYTHGSESLYLAPGGDDMIYDLGIKYSF 346

```

```

Db 179 KGAKEEL-STLHGKTYTYGPGATTIYPAGSGDDMAYDGLGKYSF 223

RESULT 14
Q9TV85 PRELIMINARY; PRT; 419 AA.
AC Q9TV85
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE A1 (EC 3.4.17.1).
GN PCP-A1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9823;
[1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99192816; PubMed=10092856;
RX Daruis S., Juge N., Marino C., Aviles F.X., Pulgarer A., Chaux J.C.,
RA Guo X.J.;
RT "Cloning, sequencing and functional expression of a cDNA encoding
RT porcine pancreatic preprocarboxypeptidase A1.";
RL Eur. J. Biochem. 259:719-725(1999).
DR EMBL; AF076222; AAD17690.1; -.
DR HSSP; P09954; IPCA.
DR MEROPS; M14.001; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CARBOXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase.
SQ SEQUENCE 419 AA; 47235 MW; 84B4CB57B714FC1 CRC64;

Query Match 27.0%; Score 516.5; DB 6; Length 419;
Best Local Similarity 33.8%; Pred. No. 2.1e-38;
Matches 132; Conservative 63; Mismatches 133; Indels 63; Gaps 12;

QY 3 LGSIAVLVPIVLCQGVHAFQSGOVLALPRISROYOVLOMTTYEIVL-WQ-PVTA 59
Db 4 LIFSIVLLGVLAKED---FVGHQVLRISYVDEAGVQVKLEDELHLDLDFWRGP 59
QY 60 DLYKKQVHFPPVNASVDNVKAKHLNVSIGIPCSVLLADVEDLI---QQQISNDTVSPRAS 116
Db 60 GFFPIDVR---VFPSIQAVKVFLEAHGIRYTIMEDVOLLDEBQOMFASQGRARTT 114
QY 117 ASY-YEQYHSLSNEIYSWIFTERHPMLTKRIHGSFEFYPLVLFVSGKEQTAKN--- 172
Db 115 STFYATVYHTLEETIDYMDLIVAEHPQVSKLQIGSSYEGRPVLYLKS---TGNNRP 170
QY 173 AIWIDCGIHAREWISPAFCIMFI----- 195
Db 171 AIWIDTGIHSREWYTAQSGWFAKRTIEDYGOQPAFTAILNDIDFLEIYTPDGFATH 230
QY 196 GHNRMKRNRSFYANNHCIGTDLNSNFYSKHMCEGASSSCSETTYGGLPESPEPEKAV 255
Db 231 SENNMKRTKRTSGSCFCVGDVDPNRNF-DAGFGAGASSNPSCEYTGKFPNSEVEVSKI 289
QY 256 ASFLRNINQIKAYISMYSQHIATFPYSTRSKSDHEELSLVASEAVRAIDKTSKNT 315
Db 290 VDEVNDHGN-ITAFISHSHSOLILYPRGYKTEPADKDELQISKAVAL-TSLYGTK 347
QY 316 YTHGSESLYLAPGGDDMIYDLGIKYSF 346
Db 348 FQYGSITTTIYQASGRTIDWTYNGIKYSFS 378

RESULT 15

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:39:53 ; Search time 28.95 Seconds
(without alignments)
303.738 Million cell updates/sec

Title: US-09-980-881-2

Perfect score: 1911

Sequence: 1 MKLCSLAVLVPIVIFCEQHV.....IKYFTSNPVEKLPLSLK 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

231628

Total number of hits satisfying chosen parameters:

231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgcn2_6/ptodata/1/1aa/5A_COMB.pep:*

2: /cgcn2_6/ptodata/1/1aa/5B_COMB.pep:*

3: /cgcn2_6/ptodata/1/1aa/5A_COMB.pep:*

4: /cgcn2_6/ptodata/1/1aa/5B_COMB.pep:*

5: /cgcn2_6/ptodata/1/1aa/PTUS_COMB.pep:*

6: /cgcn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1799.5	94.2	423	1	US-07-649-591B-3
2	1799.5	94.2	423	1	US-08-277-540-3
3	1799.5	94.2	423	1	US-08-430-787A-3
4	1799.5	94.2	423	2	US-08-869-057-2
5	640.5	33.5	404	1	US-08-869-057-2
6	639	33.4	415	2	US-08-860-882A-57
7	628.5	32.9	417	1	US-07-649-591B-7
8	628.5	32.9	417	1	US-08-277-540-7
9	628.5	32.9	417	1	US-08-430-787A-7
10	623	32.6	716	4	US-09-171-945-125
11	619.5	32.4	417	1	US-07-649-591B-6
12	619.5	32.4	417	1	US-08-277-540-6
13	619.5	32.4	417	1	US-08-430-787A-6
14	604	31.6	307	2	US-08-782-760-6
15	604	31.6	307	5	PCT-US96-00995-6
16	603.5	31.6	396	1	US-07-649-591B-4
17	603.5	31.6	396	1	US-08-277-540-4
18	603.5	31.6	396	1	US-08-430-787A-4
19	576	30.1	306	1	US-08-696-139-4
20	553	28.9	417	1	US-07-649-591B-8
21	553	28.9	417	1	US-08-277-540-8
22	553	28.9	417	1	US-08-430-787A-8
23	543	28.4	613	4	US-09-171-945-113
24	528.5	27.7	419	1	US-07-649-591B-5
25	528.5	27.7	419	1	US-08-277-540-5
26	528.5	27.7	419	1	US-08-430-787A-5
27	513	26.8	417	4	US-08-640-906-4

28	513	26.8	417	4	US-09-395-936-4	Sequence 4, App1
29	512	26.8	417	4	US-08-640-906-18	Sequence 18, App1
30	512	26.8	417	4	US-09-395-936-18	Sequence 18, App1
31	487.5	25.5	419	4	US-08-640-906-2	Sequence 2, App1
32	487.5	25.5	419	4	US-09-395-936-2	Sequence 2, App1
33	482.5	25.2	419	4	US-08-640-906-17	Sequence 17, App1
34	482.5	25.2	419	4	US-09-395-936-17	Sequence 17, App1
35	167	8.7	37	1	US-07-649-591B-1	Sequence 1, App1
36	167	8.7	37	1	US-08-277-540-1	Sequence 1, App1
37	167	8.7	37	1	US-08-430-787A-1	Sequence 1, App1
38	118.5	6.2	706	1	US-08-484-105-16	Sequence 16, App1
39	118.5	6.2	706	1	US-08-484-106-16	Sequence 16, App1
40	100.5	5.3	85	2	US-08-782-760-3	Sequence 3, App1
41	100.5	5.3	95	5	PCT-US96-00995-3	Sequence 3, App1
42	95	5.0	734	4	US-08-706-216-2	Sequence 2, App1
43	94.5	4.9	1683	3	US-08-755-587-183	Sequence 183, App
44	90.5	4.7	346	4	US-09-073-297-2	Sequence 2, App1
45	90.5	4.7	1726	2	US-08-609-049A-30	Sequence 30, App1

ALIGNMENTS

RESULT 1
US-07-649-591B-3
Sequence 3, Application US/07649591B
Patent No. 5206161
GENERAL INFORMATION:
APPLICANT: Dennis Drayna and Daniel Eaton
TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/649,591B
FILING DATE: 19910201
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/266-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-649-591B-3

Query Match 94.2%; Score 1799.5; DB 1; Length 423;
Best Local Similarity 89.6%; Pred. No. 4.2e-187;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;
OY 1 MKLCSLAVLVPIVIFCEQHVAFOSGOVLALPTSRQOVVLQNTTYYETVLVMOPTAD 60
DB 1 MKLCSLAVLVPIVIFCEQHVAFOSGOVLALPTSRQOVVLQNTTYYETVLVMOPTAD 60

QY 61 LIYKKQVHFVNASDVNDVKAHLNVSIGPCSVLLADVEDLIQOQISNDTVSPRASASY 120
DB 61 LIYKKQVHFVNASDVNDVKAHLNVSIGPCSVLLADVEDLIQOQISNDTVSPRASASY 120
QY 121 EOYHSLNEIYSWIEFTTERHPDMLTKIHIGSSFEEKYPLVLYKVSQKQOTAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFTTERHPDMLTKIHIGSSFEEKYPLVLYKVSQKQOTAKNAIWDGCI 180
QY 181 HAREWISPAFCIMFIGH-----NRMMRK 203
DB 181 HAREWISPAFCIMFIGHITQFYGIQYTNLLRLVDVFPVVPVNDGYDSMKKRRMMRK 240
QY 204 NRSFYANNHCIGTDLNSNFKHMCCEGASSSCSEYCGLYPSEPEYKAVASFLRRNI 263
DB 241 NRSFYANNHCIGTDLNSNFKHMCCEGASSSCSEYCGLYPSEPEYKAVASFLRRNI 300
QY 264 NOIKAVISMHSYQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDTKSKNTRYTHGHGSE 323
DB 301 NOIKAVISMHSYQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDTKSKNTRYTHGHGSE 360
QY 324 TLYLAPGGDDMIYDLGIRKYSFT 346
DB 361 TLYLAPGGDDMIYDLGIRKYSFT 383

RESULT 2

US-08-277-540-3

Sequence 3, Application US/08277540
Patent No. 5474901

GENERAL INFORMATION:

APPLICANT: Drayna, Dennis T., Eaton, Dan L.

TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/277, 540

FILING DATE: 19-JUL-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/167727

FILING DATE: 15-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/959944

FILING DATE: 14-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/649591

FILING DATE: 01-FEB-91

ATTORNEY/AGENT INFORMATION:

NAME: Hasak, Janet E.

REGISTRATION NUMBER: 28,616

REFERENCE/DOCKET NUMBER: 689D1C1D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1896

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 423 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-277-540-3

Query Match 94.2%; Score 1799.5; DB 1; Length 423;
Best local Similarity 89.6%; Pred No. 4.2e-187;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQOVQLONTTTEIYIWMQPVTD 60
DB 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQOVQLONTTTEIYIWMQPVTD 60
QY 61 LIYKKQVHFVNASDVNDVKAHLNVSIGPCSVLLADVEDLIQOQISNDTVSPRASASY 120
DB 61 LIYKKQVHFVNASDVNDVKAHLNVSIGPCSVLLADVEDLIQOQISNDTVSPRASASY 120
QY 121 EOYHSLNEIYSWIEFTTERHPDMLTKIHIGSSFEEKYPLVLYKVSQKQOTAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFTTERHPDMLTKIHIGSSFEEKYPLVLYKVSQKQOTAKNAIWDGCI 180
QY 181 HAREWISPAFCIMFIGH-----NRMMRK 203
DB 181 HAREWISPAFCIMFIGHITQFYGIQYTNLLRLVDVFPVVPVNDGYDSMKKRRMMRK 240
QY 204 NRSFYANNHCIGTDLNSNFKHMCCEGASSSCSEYCGLYPSEPEYKAVASFLRRNI 263
DB 241 NRSFYANNHCIGTDLNSNFKHMCCEGASSSCSEYCGLYPSEPEYKAVASFLRRNI 300
QY 264 NOIKAVISMHSYQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDTKSKNTRYTHGHGSE 323
DB 301 NOIKAVISMHSYQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDTKSKNTRYTHGHGSE 360
QY 324 TLYLAPGGDDMIYDLGIRKYSFT 346
DB 361 TLYLAPGGDDMIYDLGIRKYSFT 383

RESULT 3

US-08-430-787A-3

Sequence 3, Application US/08430787A
Patent No. 5593674

GENERAL INFORMATION:

APPLICANT: Drayna, Dennis T., Eaton, Dan L.

TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/430, 787A

FILING DATE: 27-APR-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/277, 540

FILING DATE: 19-JUL-1994

APPLICATION NUMBER: 08/167727

FILING DATE: 15-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/959944

FILING DATE: 14-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/649591

FILING DATE: 01-FEB-91

ATTORNEY/AGENT INFORMATION:

NAME: Hasak, Janet E.

REGISTRATION NUMBER: 28,616


```

COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,139
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,258
FILING DATE: 16-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-8681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 404 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-696-139-2

```

```

Query Match      33.5%; Score 640.5; DB 1; Length 404;
Best Local Similarity 37.7%; Pred. No. 4,5e-61;
Matches 136; Conservative 60; Mismatches 122; Indels 43; Gaps 6;

QY 23 FOSGOVLAALPRTSROYVQLNTTTEYLWQPVADLIYKKQYHFFVNASDVNKA 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 10 FEGEKFRVNEDENDISILHELASTROIDFWKPSDVTQIKHSTYDFVAKEDLAVED 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 83 HUNVSGIRPSVLLADVEDLIQOQISNDTVSPRASASYEQYSLNEIYEWIEFTERHDP 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 70 FLEQNLOEQVELNLRNLSVLEAFDSRV--RTTGHSEYKYNWMEITEAMTQVSENDD 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 143 MLTKIHGSSFEKPYLYLVKSGKEQTAKNAIMIDGSHAREMISPAFLMTI----- 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 127 LISRTAIGTFLGNNTIYLLK--GKBPKNPAIFMDCGFHAKMISHAFQMFVREAVLTY 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 196 -----GH-----NRMKRNSFYANNHCIGTDLNSNFYSK 225
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 186 GYESHTEFLNKLDIFYLPLVNLIDGIYTWTKRMWRKTRSTNAGTTCIGTDPNRNF-DA 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 226 HNCBEGASSSCSEYTCGLYPESEPEVKAVALFLRNINQIKAYISMSYSHIYEPYSY 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 245 GCGTTGASTDPCDEYTCGSAASEKETAKLADFIINNLSSTIKAYLTHSYSMILYPSY 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 TRSKSKDHELSLVASEAVRAIDKTSKNTRYHGHSEETLYLAPGGDDWIMYDLGIRKSF 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 305 DYKLENNALNNLAKAVKEL-ATLYGTYKYTGGATTIYPAAGSDDMAYDQIKISF 363
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 346 T 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 364 T 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 6
US-08-860-882A-57
Sequence 57, Application US/08860882A
Patent No. 5985281
GENERAL INFORMATION:
APPLICANT: TAYLORSON, CHRISTOPHER JOHN
APPLICANT: EGGELTE, HENDRIKUS JOHANNES
APPLICANT: TARRAGOMA-FIOU, ANTONIO
APPLICANT: RABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS

```

```

APPLICANT: HENNAM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
APPLICANT: DAVIES, DAVID HOW
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 COSH
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-860-882A-57

```

```

Query Match      33.4%; Score 639; DB 2; Length 415;
Best Local Similarity 35.6%; Pred. No. 6,8e-61;
Matches 135; Conservative 69; Mismatches 131; Indels 44; Gaps 7;

QY 6 LAVLPIVLFCEQHFA-FOSGOVLAALPRTSROYVQLNTTTEYLWQPVADLIYK 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 LLVLTVALASAHGGEHFEGEKFRVNEDENNIIRLASTTQIDFWKPSDVTQIKP 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 KQVHFFYNASDVNDVNAHLNVSGIPCSVLLADVEDLIQOQISNDTVSPRASASYEQYH 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 HSTYDFRVKAEDEVVENVLKQNELQYKVLISNLRNVEAFDSRV--RATGSHSEKYN 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 SLNEIYSWIEFTERHHPMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIMIDGSHARE 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 KWEITEANTQVAAENPILSRISYIGTFEGRAIYLLK--GKAQGNPAIFMDCGFHARE 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 WISPAFLMTI-----GH-----NRMKRNSF 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 WISPAFCQMFVREAVRTRYGREIQTTELLDKDIFYLPLVNLIDGIYTWTSRWRKTRST 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 YANNHCIGTDLNSNFVSKHNCBEGASSSCSEYTCGLYPESEPEVKAVALFLRNINQIK 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 238 HTGSSCIGTDPNRNF-DAGWCEIGASRNPDCDEYTCGSAASEKETAKLADFIINRLSSIK 296
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 268 AYISMHSYSHIYPPYSYTRSKSKDHELSLVASEAVRAIDKTSKNTRYHGHSEETLYL 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 297 AYLTIRHSYQMAYPIYSAKRLGNNALNNLAKAVKEL-ASLHGTYKYTGGATTIY 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 APGGDDWIMYDLGIRKSF 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 356 AAGGSDDMAYDQIRYSFT 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


Db 240 NONSTCIGTDLNRF-DVSMSSPNTNKPCLNVRGPAPESEKETKAVTNFIRSHLSIK 298
Qy 268 AYISMHSYOHIVFPYSYTRSKSDHEELSLVASEAVRAIDTSKTRTHGHGSETLYL 327
Db 299 AYTFHSYSQMLLPYGYTFKLPNHDLLKVARIAIDAL-STYETRIYIGPLASTIYK 357
Qy 328 APGGGDWITDGIKYSF 345
Db 358 TSGSLDWYDGIKHTF 375

RESULT 9
US-08-430-787A-7
Sequence 7, Application US/08430787A
Patent No. 5593674
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patln (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-7

Query Match 32.9%, Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%, Pred. No. 9.5e-60;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;

Qy 6 LAVLVPIVLFCEQHVAFQSGQVLAALPRTSRQVQVQLNLTYYEIVLPQVTAADLIYK 65
Db 7 MAVITTLAIAVH--FDREKVFVRKIQNEKHSVYLNKLQSIDLDWYDPADIHDAVN 63
Qy 66 KQVHFVNASDVNDVKALNLSGICQSVLADVELLIQOQIS-NDTVSPRASASTYEYH 124

Db 64 MTFVDFRSEKESQTIQSTLEQKHIEYELLHDLQEELEKQFVCKDELADRHS---YAKYN 120
Qy 125 SLNEIYSWIEFIERPDMILTKIHSSPEKYPFLYLKYSGEQTKAKNAIWDGCIHANE 184
Db 121 DMKIVSWTEKMLEKHEPMSRIKISITVEDNPLYLKI-GKDGERRAIFMDCGHARE 179
Qy 185 WISPAFLMEI-----GHN-----HWRNRNRSF 207
Db 180 WISPAFCQMFVQATKSYGNKIMTKLDRMNFYLPVENVGQITKSWTQDRMRKRNR 239
Qy 208 YANNHCIGTDLNNSNFKWCEGASSSCSEYGLYPESEPEKAVASFLRRINQIK 267
Db 240 NONSTCIGTDLNRF-DVSMSSPNTNKPCLNVRGPAPESEKETKAVTNFIRSHLSIK 298
Qy 268 AYISMHSYOHIVFPYSYTRSKSDHEELSLVASEAVRAIDTSKTRTHGHGSETLYL 327
Db 299 AYTFHSYSQMLLPYGYTFKLPNHDLLKVARIAIDAL-STYETRIYIGPLASTIYK 357
Qy 328 APGGGDWITDGIKYSF 345
Db 358 TSGSLDWYDGIKHTF 375

RESULT 10
US-09-171-945-125
Sequence 125, Application US/09171945
Patent No. 6277599
GENERAL INFORMATION:
APPLICANT: Emery, Stephen
APPLICANT: Copley, Clive Graham
TITLE OF INVENTION: Monoclonal Antibody to CEa, Conjugates Comprising Said
TITLE OF INVENTION: Antibody, and Their Therapeutic Use In an Adept System
FILE REFERENCE: Monoclonal Antibody to CEa
CURRENT APPLICATION NUMBER: US/09/171,945
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: GB9703103.3
PRIOR FILING DATE: 1997-02-14
PRIOR APPLICATION NUMBER: GB9609405.7
PRIOR FILING DATE: 1996-05-04
PRIOR APPLICATION NUMBER: PCT/GB97/01165
NUMBER OF SEQ ID NOS: 131
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 125
LENGTH: 716
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-125

Query Match 32.6%, Score 623; DB 4; Length 716;
Best Local Similarity 34.6%, Pred. No. 9.1e-59;
Matches 132; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

Qy 3 LCSLAVLPVLFCEQHVFA-FQSGQVLAALPRTSRQVQVQLNLTYYEIVLPQVTAADLI 61
Db 1 MLALLVLTVALASAHNGEHEFEGEKYRVNVEDENHNINIRELASITQIDFWKPDVITQ 60
Qy 62 IVKKQVHFVNASDVNDVKALNLSGICPSVLAADVEDLIQOQISNDTVSPRASASTYE 121
Db 61 IKPHSTVDFRKAEDTVEVENVLQNELQYKVLISLNKNVVEAQFDSRV--RATGHSYE 117
Qy 122 QYHSLNEIYSWIEFIERPDMILTKIHSSPEKYPFLYLKYSGEQTKAKNAIWDGCIH 181
Db 118 KYNKMETEAMTQVATBNPALISRSVIGTTFEGRATYILKV-GKAQONKPAIFMDCGFH 176
Qy 182 ARENISPAFCIMFT-----GH-----NRMRKN 204
Db 177 ARENISPAFCQMFVRAEAVRTYGREIQVTELLDKLDFVLPVNLINDGIYIWTJSRFRK 236

```

Oy 205 RSFYANNICIGDILNSNFVSKMCEGASSSCSEFYCGLYPESPEYKAVASFLPRNTN 264
    || : ||||| || || ||| : ||||| || | ||: || :
Db 237 RSTHGSSCIGTDPPRNF-DAGWCITGASRNCDETTGCPAAESKETKALDFTRNKIS 295
Oy 265 QIKAYISMHSYQHIIVFPYSTYRSKCKHHEELSLVASEAVRAIDRTSKNTRYTHGSGST 324
    ||||:||||| :||| || :|| : || : || : || : || :
Db 296 SIKAVLTJHSYSQMMIYPSYAYKRGENNAELNALAKATVKRL-ASLHGTXYTGPGATT 354
Oy 325 LYLAFGGDDMTYDGLIRYSFT 346
    : | : || || || || || ||
Db 355 IYPSAGTSKDMAYDGLIRYSFT 376

RESULT 11
US-07-649-591B-6
: Sequence 6, Application US/07649591B
: Patent No. 5206161
: GENERAL INFORMATION:
: APPLICANT: Dennis Drayna and Daniel Eaton
: TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/649,591B
: FILING DATE: 19910201
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 889
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/266-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 417 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: US-07-649-591B-6

Query Match 32.4%; Score 619.5; DB 1; Length 417;
Best Local Similarity 34.1%; Pred. No. 9e-59;
Matches 130; Conservative 76; Mismatches 128; Indels 47; Gaps 8;

Oy 6 LALVLPYILFCQHFVA--FOSGOYLAALPFTSKQVOYLOLMTTYEILVLMQPYADLI 62
    : ||| : | | | | | : | : | : | : | : | : | :
Db 1 MRLILPVGILATTTAIAAPRFDEKVFVRVKKPODEKQADIIKDLARTNELDFYPGATHIV 60
Oy 63 YKKKQVHFVNASDVNKAHLNVSIGIPCSVLLADVEDLIOQOIS-NDTVSPRASASYE 121
    | | | : : : : : : : : : | : | : | : | : | :
Db 61 AANMMVDRVSKESQALIOSALDQKKMAYEILIHDLQETEKQFVKEDIPGRHS--YA 117
Oy 122 QYHSNLNETSWIEFTTERHPDMLTKIHIGSSPEKYPLVLYKVSGRQETAKNAIWDGCIH 181
    :| : | : | : | : | : | : | : | : | : | : | : | :
Db 118 KYNNEKEIVAMTEKMKDKPEKVSYHKGISTYEDNPPLYLKI-GEKNEKRKAIFMDCGIH 176
Oy 182 AREMISPAFLCMTF-----GH-----NRMWRKN 204

```

```

Db      177 AREWSPAFECQFFVYQARKTYGRNKKIMTKLDRBMFYLLPVPNVGCIYIMSTWKKRMRKN 236
QY      205 RSEFYANNHCIGTDLNSNFVSKHWCEGASSSSCSETCGYGLYPESEDEPKAVASFLRRNIN 264
Db      237 RSKNONSNCICIGDLNRNF-NASWMSIPIPTNDPCADNRCGSAPESEKETATVNFIRSHDN 295
QY      265 OIKAVIIMSHSVSOHIVPFPSTYRSKSKHKEELSLVASEAVRAIDTSTKRTYTHHGSGST 324
Db      236 EKVYITHHSISQMLFFPGYTSKILPNNHDLAKVAKIGTDVL-STRETRYIIGPIEST 354
QY      325 IYLAPGGGDWYDLGIRYSEF 345
Db      355 IYPIGSSLDMAWYDLGIRHTEF 375

RESULT 12
US-08-277-540-6
; Sequence 6, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5 25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28, 616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-277-540-6

Query Match 32.4%; Score 619.5; DB 1; Length 417;
Best Local Similarity 34.1%; Pred. No. 9e-59;
Matches 130; Conservative 76; Mismatches 128; Indels 47; Gaps 8;

6 IAVLPVIVLFCQHWFA--FOSGOVLALPPTSROVOVLNLTYYEIVLMOPVTADLI 62
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MLULIPVGLIATTLTALAPVPRFDEKVFRRKPODEKADITIKIDAKTNELDFFWPGATHHV 60

```

QY 63 VKKQVHEFYVNASDVNDVNAHLNVSGIPCSVLLADVEDLIQOIS-NDTVSPRASASYE 121
 Db 61 AAMMMVDFRVSKEQALSDONKMHYEILIHDLQEEIEKQFVKEDIPGRHS--YA 117
 QY 122 QYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLVAKSGEQTAKNAIMWDCGIH 181
 Db 118 KYNMKEIYAMTEKMDKPEWYSRIKIGSTVEDNPLYVAKI-GEKNERKALFMDCGIH 176
 QY 182 AREWISPAFCIMFI-----GH-----NRMWRKN 204
 Db 177 AREWSPAFQCMFVYQATKTYGRNKMIMTKLLDRMNFYILFVENVGCIYIMSWTKNRMWRKN 236
 QY 205 RSFYANNHCIGTDLNSNEVSKHWCCEGASSCSEYTGILPESEPEVAVASFLLRNIN 264
 Db 237 RSKNNSKICIGDILNRF-NASMNSIPNTNDPCADNYRGSAPESKEKAVTNFRSHLN 295
 QY 265 QIKAVISMHSYQOHLVFPYSYTRSKSKDHEELSLVASEAVRAIDKSNTRYTHGCHSET 324
 Db 296 EIKVYITFHSYQMLLFPYGYTSKLPNNHEDLAKVAKIGTDVL-STRYETRIYIGPIEST 354
 QY 325 LYIAPGGDDWYIDGIRYSF 345
 Db 355 IYPIGSSSLDMAYDIDGIRKHTF 375

RESULT 13

US-08-430-787A-6
 ; Sequence 6, Application US/08430787A
 ; Patent No. 5593674
 ; GENERAL INFORMATION:
 ; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
 ; TITLE OF INVENTION: No. 5593674e1 Plasma Carboxypeptidase
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 460 Point San Bruno Blvd
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: patin (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/430,787A
 ; FILING DATE: 27-APR-1995
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/277,540
 ; FILING DATE: 19-JUL-1994
 ; APPLICATION NUMBER: 08/167727
 ; FILING DATE: 15-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/959944
 ; FILING DATE: 14-OCT-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/649591
 ; FILING DATE: 01-FEB-91
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hasak, Janet E.
 ; REGISTRATION NUMBER: 28,616
 ; REFERENCE/DOCKET NUMBER: 689D1C1D1
 ; TELEPHONE: 415/225-1896
 ; TELEFAX: 415/952-9881
 ; TELEX: 910/371-7168
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 417 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear

US-08-430-787A-6

Query Match 32.4%; Score 619.5; DB 1; Length 417;
 Best Local Similarity 34.1%; Pred. No. 9e-59;
 Matches 130; Conservative 76; Mismatches 128; Indels 47; Gaps 8;

QY 6 LAVLPVLFCEQHVFA---FQSGVIALPRTSRQVQVQMLTTYETVLMQPTADLI 62
 Db 1 MRLILFVGLIATTLAIPVRFREKVFYRKPODERQADIIKDLATNELDFWYPCATHV 60
 QY 63 VKKQVHEFYVNASDVNDVNAHLNVSGIPCSVLLADVEDLIQOIS-NDTVSPRASASYE 121
 Db 61 AAMMMVDFRVSKEQALSDONKMHYEILIHDLQEEIEKQFVKEDIPGRHS--YA 117
 QY 122 QYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLVAKSGEQTAKNAIMWDCGIH 181
 Db 118 KYNMKEIYAMTEKMDKPEWYSRIKIGSTVEDNPLYVAKI-GEKNERKALFMDCGIH 176
 QY 182 AREWISPAFCIMFI-----GH-----NRMWRKN 204
 Db 177 AREWSPAFQCMFVYQATKTYGRNKMIMTKLLDRMNFYILFVENVGCIYIMSWTKNRMWRKN 236
 QY 205 RSFYANNHCIGTDLNSNEVSKHWCCEGASSCSEYTGILPESEPEVAVASFLLRNIN 264
 Db 237 RSKNNSKICIGDILNRF-NASMNSIPNTNDPCADNYRGSAPESKEKAVTNFRSHLN 295
 QY 265 QIKAVISMHSYQOHLVFPYSYTRSKSKDHEELSLVASEAVRAIDKSNTRYTHGCHSET 324
 Db 296 EIKVYITFHSYQMLLFPYGYTSKLPNNHEDLAKVAKIGTDVL-STRYETRIYIGPIEST 354
 QY 325 LYIAPGGDDWYIDGIRYSF 345
 Db 355 IYPIGSSSLDMAYDIDGIRKHTF 375

RESULT 14

US-08-782-760-6
 ; Sequence 6, Application US/08782760
 ; Patent No. 5948668
 ; GENERAL INFORMATION:
 ; APPLICANT: Hartman, Jacob
 ; APPLICANT: Fulga, Netta
 ; APPLICANT: Wendelovitch, Simona
 ; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/782,760
 ; FILING DATE: 13-JAN-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/378,233
 ; FILING DATE: 25-JAN-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 0336/43847
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:38:53 ; Search time 65 Seconds

(without alignments)
615.178 Million cell updates/sec

Title: US-09-980-881-2

Perfect score: 1911
Sequence: 1 MKLCSLAVLPVLFCEQHV.....IKYFTSNPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

1: A.Geneseq_032802:*
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1911	100.0	360	21	AA11457 Human brain carbox
2	1799.5	94.2	423	17	AA190293 Human plasma carbo
3	1799.5	94.2	423	18	AA114733 Human plasma carbo
4	1799.5	94.2	423	20	AA192270 Human plasma carbo
5	1794.5	93.9	423	14	AA136273 Human plasma carbo
6	1785	93.4	338	21	AA11458 Human brain carbox
7	1325	69.3	246	21	AA11459 Human brain carbox
8	803.5	42.0	211	21	AA158129 Lung cancer associ
9	640.5	33.5	404	16	AA175131 Porcine Tyr-His-Me
10	639	33.4	415	17	AA106172 Human pancreatic c
11	639	33.4	415	19	AA174476 Human pancreatic c

12	631.5	33.0	424	17	AA106175
13	629.5	32.9	424	22	AA165652
14	623	32.6	716	18	AA141414
15	604	31.6	307	17	AA100602
16	604	31.6	437	22	AA100502
17	576	30.1	306	16	AA175132
18	576	30.1	306	22	AA104477
19	563	29.5	354	22	AA165650
20	563	29.5	354	22	AA165656
21	563	29.5	354	22	AA174682
22	563	29.5	374	22	AA165654
23	563	29.5	374	22	AA165655
24	561	29.4	329	17	AA106174
25	561	29.4	349	17	AA106173
26	554	29.0	349	17	AA106181
27	554	28.9	349	18	AA113749
28	553	28.9	349	17	AA106182
29	553	28.9	349	18	AA113750
30	550	28.8	349	18	AA113753
31	548	28.7	349	18	AA113752
32	548	28.7	349	18	AA113758
33	547	28.6	349	18	AA113751
34	546	28.6	349	18	AA113753
35	546	28.6	399	22	AA101664
36	545	28.5	311	22	AA165653
37	545	28.5	349	18	AA113760
38	545	28.5	349	18	AA113761
39	545	28.5	349	18	AA113762
40	543	28.4	349	18	AA113757
41	543	28.4	349	18	AA113765
42	543	28.4	613	19	AA141412
43	543	28.4	613	19	AA174481
44	542	28.4	349	18	AA113759
45	540	28.3	349	18	AA113764

ALIGNMENTS

RESULT 1	AA11457	standard; Protein: 360 AA.
ID	AA11457	
AC	AA11457	
XX		
DT	01-MAR-2001	(first entry)
XX		
DE	Human brain carboxypeptidase B protein.	
XX		
KW	Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase; treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine; neuroprotective; antialzheimers; neurotrophic; neuroprotective;	
KW	hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;	
KW	Down's syndrome; head trauma.	
OS	Homo sapiens.	
XX		
PN	WO200066717-A1.	
XX		
PD	09-NOV-2000.	
XX		
PF	01-MAY-2000; 2000WO-JP02878.	
XX		
PR	30-APR-1999; 99JP-0125169.	
XX		
PA	(MATS/) MATSUMOTO A.	
XX		
PI	Matsumoto A;	
XX		
DR	WPI; 2000-687534/67.	
DR	N-PSDB; AAC1962.	
XX		
PT	Human brain carboxypeptidase B isolated from the hippocampus useful for	

ProHCPB with PeB
Bothrops jararaca C
PreproHCPB-linker-
Rat carboxypeptida
Human carboxypepti
Porcine carboxypep
Porcine carboxypep
Human secreted met
Human secreted met
Human protease and
Human secreted met
Human secreted met
Mature HCPB with P
Modified HCPB (H1s)6
Modified HCPB (D25)
Modified HCPB (D25)
Carboxypeptidase B
Carboxypeptidase B
Carboxypeptidase B
Carboxypeptidase B
Carboxypeptidase B
Carboxypeptidase B
Novel human protea
Human carboxypepti
Carboxypeptidase B
Carboxypeptidase B
Carboxypeptidase B
Carboxypeptidase B
Carboxypeptidase B
Carboxypeptidase B
Human pancreatic c
Carboxypeptidase B
Carboxypeptidase B


```
DB 241 nrsfyannhcgicgldlnnfaskhwcceagssscsetyqglypesepevkavasflrrni 300
QY 264 NQIKAYISMHSYSQHIYFPYSYTRSKSKDHELSLVAEAVRAIDKTSKNTRYTHGHGSE 323
DB 301 nqikayismhsyqghlyfpystrskskdheelslvaseavratelctskotrythghgse 360
QY 324 TLYLAPGGGDDWLYDLGIKYSFT 346
DB 361 tlylapgggddwlydlgikysft 383

RESULT 3
AAW14733
ID AAW14733 standard; Protein; 423 AA.
AC AAW14733;
XX
DT 08-MAY-1997 (first entry)
DE Human plasma carboxypeptidase B.
KM Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
XX plasma; plasmalogen.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT Protein /note= "Signal peptide"
FT /note= "Mature PCPB"
FT Cleavage-site 114
FT /note= "Potential clip site for activation of PCPB
as a carboxypeptidase"
FT Active-site 181
FT /note= "Expected to be involved in catalytic activity"
FT Binding-site 183
FT /note= "Expected to be involved in substrate binding"
FT Active-site 184
FT /note= "Expected to be involved in catalytic activity"
FT Binding-site 256
FT /note= "Expected to be involved in substrate binding"
FT Binding-site 257
FT /note= "Expected to be involved in substrate binding"
FT Active-site 310
FT /note= "Expected to be involved in catalytic activity"
FT Binding-site 312
FT /note= "Expected to be involved in substrate binding"
FT Active-site 369
FT /note= "Expected to be involved in catalytic activity"
FT Binding-site 370
FT /note= "Expected to be involved in substrate
binding, thought to determine specificity
of PCPB as a carboxypeptidase B"
FT Binding-site 371
FT /note= "Expected to be involved in substrate binding"
FT Binding-site 394
FT /note= "Expected to be involved in substrate binding"
XX
PN US5593674-A.
XX
PD 14-JAN-1997.
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
PR 27-APR-1995; 95US-0430787.
XX
XX (GETH ) GENENTECH INC.
```

```
PI Drayna DT, Eaton DL;
XX
DR WPI: 1997-099413/09.
DR N-PSDB; AAT62846.
XX
PT Using human plasma carboxypeptidase B in blood coagulation - 1s
PT functionally related to carboxypeptidase A and pancreas
PT carboxypeptidase B
XX
PS Example 2; Column 37-42; 39pp; English.
XX
CC This sequence represents human plasma carboxypeptidase B (PCPB) which
CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
CC purified from human plasma or by transformed cell culture by
CC extraction using plasmalogen bound to a solid phase.
XX
SQ Sequence 423 AA;

Query Match 94.2%; Score 1799.5; DB 18; Length 423;
Best Local Similarity 89.6%; Pred. No. 3.7e-176;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVLPFCEQHVPFQSGOYLAALPRTSRQVOYLNTTTEIVLMQPTAD 60
DB 1 mklcslavlpvlpfceqhvafqsgylaalprrtsrqvylqnltyelwqpvfad 60
QY 61 LIYKKQVHFVNASDNDNVKAHLNVSGIPCSVLADVEDLIQOQISMDTSPRASVY 120
DB 61 liykkqvhtfvnasdndnvkahlnvsgipcsvlladvedliqqisndtvsprassyy 120
QY 121 EOYSLNEIYSWIEFITERHHPDMLTKIHGSSFEKYPYLVKSGEQTAKNAIIMCGI 180
DB 121 eqyslneiyswiefiterhpdmltkihgssfekypylvksgeqtaknalwidqgi 180
QY 181 HAREWISPAFLWFTGH-----NRMAR 203
DB 181 harewispaflwftghlftgylgyltnllrvdlympvrvndgydyswkknmrwx 240
QY 204 NRSFYANNHCIGTDLNLSFVSKHWCCEGASSSCSEYTCGYPSEPEVKAVASFLRRNI 263
DB 241 nrsfyannhcgicgldlnnfaskhwcceagssscsetyqglypesepevkavasflrrni 300
QY 264 NQIKAYISMHSYSQHIYFPYSYTRSKSKDHELSLVAEAVRAIDKTSKNTRYTHGHGSE 323
DB 301 nqikayismhsyqghlyfpystrskskdheelslvaseavratelctskotrythghgse 360
QY 324 TLYLAPGGGDDWLYDLGIKYSFT 346
DB 361 tlylapgggddwlydlgikysft 383

RESULT 4
AAW92270
ID AAW92270 standard; Protein; 423 AA.
AC AAW92270;
XX
DT 28-APR-1999 (first entry)
DE Human plasma carboxypeptidase B (PCPB) thr147.
XX
KM Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW polymorphism detection; thrombotic disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT Protein /note= "Signal peptide"
FT /note= "Mature PCPBthr147"
```


CC between the deduced preproPCPB amino acid sequence and pro-human
 CC mast cell carboxypeptidase A and between preproPCPB and prepro-rat
 CC carboxypeptidase B. Human PCPB has the same substrate binding sites
 CC as, and shares 6 cysteine residues which form 3 intramolecular
 CC disulphide bonds with. bovine and rat carboxypeptidase B. The
 CC presence of the same amino acid (asp at position 348 of the mature
 CC protein) at the region in carboxypeptidases that determines
 CC substrate specificity, suggests that PCPB represents a plasma-
 CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
 CC tPA of plasminogen to plasmin in the presence of fibrinogen.

XX Sequence 423 AA:

Query Match 93.9%; Score 1794.5; DB 14; Length 423;
 Best Local Similarity 89.3%; Pred. No. 1.2e-175;
 Matches 342; Conservative 1; Mismatches 3; Indels 37; Gaps 1;

QY 1 MKICSLAVLVPIVLFCEQHFAFGSGVLAALPRTSRQVQLNLTFTYEVLMQPTAD 60
 DB 1 MKICSLAVLVPIVLFCEQHFAFGSGVLAALPRTSRQVQLNLTFTYEVLMQPTAD 60
 QY 61 LIYKKQVHFVNASVDVNAKALNVSGLPCSVLADVEDLIQQQISNDTSPRASASY 120
 DB 61 LIYKKQVHFVNASVDVNAKALNVSGLPCSVLADVEDLIQQQISNDTSPRASASY 120
 QY 121 EGYHSINETYSWIEFTEERRHDMLEKIHGSEFEKYPVLYLVKSGKEQTAKNAIWDGCI 180
 DB 121 EGYHSINETYSWIEFTEERRHDMLEKIHGSEFEKYPVLYLVKSGKEQTAKNAIWDGCI 180
 QY 181 HAREWISPAFCIMFIGH-----NRMRK 203
 DB 181 HAREWISPAFCIMFIGH-----NRMRK 203
 QY 204 NRSFYANNHCIGDNLNSNFYSKHMCEGASSSCSEFYCGLYPESPEVKAASFLEARNI 263
 DB 241 NRSFYANNHCIGDNLNSNFYSKHMCEGASSSCSEFYCGLYPESPEVKAASFLEARNI 263
 QY 264 NQIKAYISMHSYSQHIIVFPYSYTRRSKSDHEELSLVASEAVRAIDKTSKTRTTHGSGS 323
 DB 301 NQIKAYISMHSYSQHIIVFPYSYTRRSKSDHEELSLVASEAVRAIDKTSKTRTTHGSGS 360
 QY 324 TLYLAPGGDDWITLDGIRKYSFT 346
 DB 361 TLYLAPGGDDWITLDGIRKYSFT 383

RESULT 6

AAB11458
 ID AAB11458 standard; Protein; 338 AA.

XX AAB11458;
 XX
 DT 01-MAR-2001 (first entry)

DE Human brain carboxypeptidase B protein SEQ ID NO 3.
 XX
 XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimer's; neurotrophic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.

XX Homo sapiens.

XX WO200066717-A1.

XX 09-NOV-2000.

XX 01-MAY-2000; 2000WO-JP02878.

XX 30-APR-1999; 99JP-0125169.

PA (MATS/) MATSUMOTO A.

XX Matsumoto A;

XX WPI; 2000-687534/67.

XX Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -

PS Disclosure; Page 71-73; 84pp; Japanese.

CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotrophic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.

XX Sequence 338 AA:

Query Match 93.4%; Score 1785; DB 21; Length 338;
 Best Local Similarity 99.1%; Pred. No. 7.9e-175;
 Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 23 FQSGQVLAALPRTSRQVQLNLTFTYEVLMQPTADLIYKKQVHFVNASVDVNAKA 82
 DB 1 FQSGQVLAALPRTSRQVQLNLTFTYEVLMQPTADLIYKKQVHFVNASVDVNAKA 60
 QY 83 HLNVSGLPCSVLADVEDLIQQQISNDTSPRASASYEGYHSINETYSWIEFTEERRHPD 142
 DB 61 HLNVSGLPCSVLADVEDLIQQQISNDTSPRASASYEGYHSINETYSWIEFTEERRHPD 120
 QY 143 MLTKIHGSEFEKYPVLYLVKSGKEQTAKNAIWDGCIHAREWISPAFCIMFIGHNRMR 202
 DB 121 MLTKIHGSEFEKYPVLYLVKSGKEQTAKNAIWDGCIHAREWISPAFCIMFIGHNRMR 180
 QY 203 KNRSEFYANNHCIGDNLNSNFYSKHMCEGASSSCSEFYCGLYPESPEVKAASFLEARNI 262
 DB 181 KNRSEFYANNHCIGDNLNSNFYSKHMCEGASSSCSEFYCGLYPESPEVKAASFLEARNI 240
 QY 263 INQIKAYISMHSYSQHIIVFPYSYTRRSKSDHEELSLVASEAVRAIDKTSKTRTTHGSGS 322
 DB 241 INQIKAYISMHSYSQHIIVFPYSYTRRSKSDHEELSLVASEAVRAIDKTSKTRTTHGSGS 300
 QY 323 ETLYLAPGGDDWITLDGIRKYSFTSNPVEKLLPLSLK 360
 DB 301 ETLYLAPGGDDWITLDGIRKYSFTSNPVEKLLPLSLK 338

RESULT 7

AAB11459
 ID AAB11459 standard; Protein; 246 AA.

XX AAB11459;

XX 01-MAR-2001 (first entry)

DE Human brain carboxypeptidase B protein SEQ ID NO 4.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimer's; neurotrophic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.

XX Homo sapiens.

XX WO200066717-A1.

```

XX 09-NOV-2000.
PD
XX
XX 01-MAY-2000; 2000WO-JP02878.
PF
XX
XX 30-APR-1999; 99JP-0125169.
PR
XX (MATSU/) MATSUMOTO A.
PA
XX
XX Matsumoto A;
PI
XX
XX WPI: 2000-687534/67.
DR
XX
XX Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
XX
XX Disclosure: Page 74-75; 84pp; Japanese.
PS
XX
XX This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC neurotropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
XX
XX Sequence 246 AA;
SQ
XX
XX
XX Query Match 69.3%; Score 1325; DB 21; Length 246;
XX Best Local Similarity 98.8%; Pred. No. 1e-127;
XX Matches 243; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX
XX 115 ASASYEEOHSLNETYSEFTTERRHPDLTKIHGSSPEKYLVLVYSGKEGTAKNAI 174
DB 1 aasayyeqhslnetyselftttrhpdltkighsspekylvlvysgkegtaknai 60
QY 175 WDCGTHAREWISPAFCIMFICGHNMRKNSFYANNHCIGTDLNSNFYSKIMCEGASS 234
DB 61 wdcgtharewispafclmfcighnmrknsfyannhcgldlnrnfaskmcegass 120
QY 235 SSCSEYCGIYPSEPEYKAVASFLRRNINQKATISMHSYSGHIYFPRSTRSKDHE 294
DB 121 sscsecycgilypsepevkavasflrrlnqkayismhsyqghivfpystrskskdhe 180
QY 295 ELSLVASEAVRAIDKTSKNTRYTHGSEETLYLAPGGDDWIYDGIKYSFSTSNPVEKL 354
DB 181 elslvasaevrailektsknttrythgsetlylapggddwlydgiyksfstsnpvekl 240
QY 355 LPLSLK 360
DB 241 lplslk 246
XX
XX
XX RESULT 8
XX AAB58129
XX ID AAB58129 standard; Protein; 211 AA.
XX AC AAB58129;
XX
XX 14-MAR-2001 (first entry)
XX
XX Lung cancer associated polypeptide sequence SEQ ID 467.
XX
XX Human; lung cancer associated protein; neuroprotective; cytostatic;
XX cardioactive; immunomodulatory; muscular active; vulnery;
XX gastrointestinal; nephrotropic; antinflective; gynecological;
XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX proliferative disorder; wound healing; infectious disease.
XX

```

```

OS Homo sapiens.
XX
XX WO200055180-A2.
PN
XX
XX 21-SEP-2000.
PD
XX
XX
XX 08-MAR-2000; 2000WO-US05918.
PF
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM;
PI
XX
XX WPI: 2000-587514/55.
DR
XX N-PSDB; AAF18005.
DR
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX
XX Claim 11; Page 944-945; 1425pp; English.
PS
XX
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective, cytostatic, cardioactive,
CC immunomodulatory, muscular active, general; vulnery; gastrointestinal
CC general; nephrotropic; antinflective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterization of the polynucleotide and protein
CC sequences.
XX
XX
XX Sequence 211 AA;
SQ
XX
XX
XX Query Match 42.0%; Score 803.5; DB 21; Length 211;
XX Best Local Similarity 88.8%; Pred. No. 3.8e-74;
XX Matches 151; Conservative 2; Mismatches 8; Indels 9; Gaps 2;
XX
XX
XX 185 WISPAFCIMFT-----GHNRRMRKNSFYANNHCIGTDLNSNFYSKIMCEGASSS 236
DB 3 wis-pa-fclmft-----ghnrmrknsfyannhcgldlnrnfaskmcegasss 61
QY 237 CSEFYCGIYPSEPEYKAVASFLRRNINQKATISMHSYSGHIYFPRSTRSKDHEEL 296
DB 62 csefycgilypsepevkavasflrrlnqkayismhsyqghivfpystrskskdheel 121
QY 297 SLVASEAVRAIDKTSKNTRYTHGSEETLYLAPGGDDWIYDGIKYSFST 346
DB 122 slvasaevrailektsknttrythgsetlylapggddwlydgiyksfst 171
XX
XX
XX RESULT 9
XX AAR75131
XX ID AAR75131 standard; Protein; 404 AA.
XX AC AAR75131;
XX
XX 13-MAR-1996 (first entry)
XX
XX Porcine Tyr-His-Met Procarboxypeptidase B.
XX

```



```

OY 6 LAVLPVILFCEQHVFA-FQSGQVLAALPRTSRQOVQVQNTTTEIYIMQPTADLIYK 64
DB 2 LVLVTVaLaSaHghgeKvfrvnedenhlllrelastqldfwtkpsvtqikp 61
OY 65 KQOVHFEVNASDVNDVKAHLNWSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEYOH 124
DB 62 hstvftrvkaedltvtenvllkqnelqykvllsnlnrvaeqfdrsv---ratghsyekyn 118
OY 125 SINEIYSWIEFTTERHPMLTKIHGSSFEKYPPLVYLKVSQKQAKNAIWDGCIHARE 184
DB 119 kwetleawtqvatentpalistsvgttfeqraiyllkv-gkagqnpkpatfmdcgfhare 177
OY 185 WISPAFCIMFTI-----GH-----NRMNRKNSRF 207
DB 178 wlsptacqfwfveavrtvgtrelqvtellldkfylvlpvlnldgylytwtkstfrwtktst 237
OY 208 YANNHCIGTDLNSNFVSKHMCCEGASSSCSETCGLYPESEPEYKAVASFLRRNINOIK 267
DB 238 htgsacigtprnrf-dagwcelgastnrcpdcetycpaaseseketaldflrnklsak 296
OY 268 AYISMSHSYQHIYFEPSTYRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGSEETLYL 327
DB 297 aylthysqmmilypysyayklgennaelnalakatvkel-ashgkykytgyppattilyp 355
OY 328 APGGDDMIYDLGIKYSFT 346
DB 356 aagsddwadydqlytsft 374

RESULT 11
AAW74476
ID AAW74476 standard; Protein: 415 AA.
XX
AC AAW74476;
DT 20-NOV-1998 (first entry)
DE Human pancreatic carboxypeptidase protein sequence.
XX
KW Human; pancreatic carboxypeptidase B; insulin; protein sequencing;
  produg therapy.
XX
OS Homo sapiens.
XX
PM W09835988-A1.
PD 20-AUG-1998.
PE 10-FEB-1998; 98WO-GB00415.
XX
PR 29-OCT-1997; 97GB-0022727.
PR 14-FEB-1997; 97GB-0003104.
PR 18-OCT-1997; 97GB-0022003.
XX
PA (ZENE ) ZENEPA (ZENE ) ZENECA LTD.
XX
PI Edge MD;
XX
DR WPI; 1998-467168/40.
DR N-PSDB; AAV41795.
XX
XX
PT New modified pro-domain of carboxy-peptidase B - enhances expression
PT of co-expressed proteins for production of recombinant
PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT enzyme produg therapy
XX
XX
PS Example 1; Page 54-55; 83pp; English.
XX
XX
CC The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC modified pro-domain of CPB on a separate gene to enhance recombinant
CC expression. This process can be used to produce recombinant CPB in
CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
CC insulin production and protein sequencing, while its fusions with

```

```

CC antibodies are useful in antibody-directed enzyme produg therapy. The
CC modified pro-domain provide increased yield of recombinant CPB, possibly
CC by protecting the C-terminus against enzymatic degradation or by
CC increasing intracellular trafficking.
SQ Sequence 415 AA:

Query Match 33.4%; Score 639; DB 19; Length 415;
Best Local Similarity 35.6%; Pred. No. 9.7e-57;
Matches 135; Conservative 69; Mismatches 131; Indels 44; Gaps 7;

OY 6 LAVLPVILFCEQHVFA-FQSGQVLAALPRTSRQOVQVQNTTTEIYIMQPTADLIYK 64
DB 2 LVLVTVaLaSaHghgeKvfrvnedenhlllrelastqldfwtkpsvtqikp 61
OY 65 KQOVHFEVNASDVNDVKAHLNWSGIPCSVLLADVEDLIQOOISNDTVSPRASASYEYOH 124
DB 62 hstvftrvkaedltvtenvllkqnelqykvllsnlnrvaeqfdrsv---ratghsyekyn 118
OY 125 SINEIYSWIEFTTERHPMLTKIHGSSFEKYPPLVYLKVSQKQAKNAIWDGCIHARE 184
DB 119 kwetleawtqvatentpalistsvgttfeqraiyllkv-gkagqnpkpatfmdcgfhare 177
OY 185 WISPAFCIMFTI-----GH-----NRMNRKNSRF 207
DB 178 wlsptacqfwfveavrtvgtrelqvtellldkfylvlpvlnldgylytwtkstfrwtktst 237
OY 208 YANNHCIGTDLNSNFVSKHMCCEGASSSCSETCGLYPESEPEYKAVASFLRRNINOIK 267
DB 238 htgsacigtprnrf-dagwcelgastnrcpdcetycpaaseseketaldflrnklsak 296
OY 268 AYISMSHSYQHIYFEPSTYRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGSEETLYL 327
DB 297 aylthysqmmilypysyayklgennaelnalakatvkel-ashgkykytgyppattilyp 355
OY 328 APGGDDMIYDLGIKYSFT 346
DB 356 aagsddwadydqlytsft 374

RESULT 12
AAW06175
ID AAW06175 standard; Protein: 424 AA.
XX
AC AAW06175;
DT 17-FEB-1997 (first entry)
DE PROHCPB with PeLB leader sequence.
XX
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPBT;
KW mustard-ribonucleotide; antibody directed enzyme produg therapy;
KW anti-neoplastic; produg; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Ed; Fab'12; PeLB; leader; human carboxypeptidase B.
XX
OS Synthetic.
XX
XX
FH Key
FH Peptide 1..22 Location/Qualifiers
FT /label= PeLB_leader_sequence
FT Peptide 23..117
FT /label= pro_sequence
FT Protein 118..424
FT /label= mature_HCPB
XX
PN W09620011-A1.
XX
XX
PD 04-JUL-1996.
XX
PF 21-DEC-1995; 95WO-GB02991.
XX

```


PR 16-AUG-1995; 95GB-0016810.
 PR 23-DEC-1994; 94GB-0026192.
 XX
 PA (ZENE) ZENECA LTD.
 XX
 PI Blakey DC, Boyle FT, Davies DR, Eggelte HJ, Heaton DW,
 PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
 PI Tarragona-Fiol A, Taylorson CJ;
 DR WPI: 1996-321650/32.
 DR N-PSDB; AAT42506.
 XX
 PT Two component system for anti-tumour therapy - comprising targeting
 PT moiety linked to mutated enzyme which can transform an
 PT anti-neoplastic produg
 XX
 PS Reference Example 18; Page 140-142; 182pp; English.
 CC A two-component system for anti-tumour therapy comprises a targeting
 CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
 CC produg. The system is based on antibody directed enzyme produg therapy
 CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
 CC pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
 CC Alternatively a modified human pancreatic carboxypeptidase B (HCBP) can
 CC be used. The present sequence is that of proKCPB with a Pelt leader
 CC sequence encoded by pCIT1738.
 CC
 XX
 SQ Sequence 424 AA;
 XX
 Query Match 33.0%; Score 631.5; DB 17; Length 424;
 Best Local Similarity 35.7%; Pred. No. 6e-56;
 Matches 129; Conservative 68; Mismatches 121; Indels 43; Gaps 6;
 XX
 QY 23 FOSGQVLAALPRTSRQVQLONTTTEIYVLMOPVADLYKKQVFEVNASVDNKA 82
 DB 29 feqekvrrvvedenhhilirelastqgidvkrpdsvtqlkphstvdfrvaedltvven 88
 QY 83 HLNVSIGPCSVLADVEDLIQOQISNDTSPRASASYEYQHSNLETYSWIEFTERRHPD 142
 DB 89 vlkqnelqykvllshlnvveaqdsrv---ratghayekynkwellvawtqvatempa 145
 QY 143 MLRTIHIGSSPEKRYPLVAVSGKEQTAKNAIWDGCIHAREWISPAFLMFT----- 195
 DB 146 llsrsvlgltfegrailkv-gkaagqkallmdcgfhawelspafcgwfvreaavrtly 204
 QY 196 -----GH-----NRMWRKRNSFYANNHCIGTDLSNFFYSK 225
 DB 205 grelqvteilldkidfyilpvlndgyvyswksrmwrtkrsvnaegstcigtprnrf-daa 263
 QY 226 HMCESGASSSSCSEYTCGLYPESEPEVKAVASFLRRINQIKAYISMHSYSQHIYFPYSY 285
 DB 264 gwceigsarmpcdelycgpaeseketkaladfirnklsikaylthhsyqgmmllypsys 323
 QY 286 TRSKSKHHELSIVASRAVAIDKTSKNTRTYTHGSETLYLAGGGDDWYIDGIKYSF 345
 DB 324 aylkgennaelnalakatvkel-aslngtkylytgpattlypaagsddwagdgqlyrsf 382
 QY 346 T 346
 DB 383 t 383
 XX
 RESULT 13
 AAG6562
 ID AAG6562 standard; Protein; 324 AA.
 XX
 AC AAG6562;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Bothops jararaca carboxypeptidase homologue protein.
 XX

KW secreted metallocarboxypeptidase; immunomodulatory;
 KW antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
 KW osteopathic; antiarthritic; cerebroprotective; vasotropic; vulnerary;
 KW antibacterial; antiviral; antifungal; antiinflammatory; antidiabetic;
 KW antineumatic; cytostatic; antiallergic; haemostatic; infection;
 KW cancer.
 XX Bothops jararaca.
 OS
 PN WO200157265-A1.
 PN
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US03783.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 29-SEP-2000; 2000US-0676135.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Boyle BJ, Mize NK, Arterburn MC, Yeung G, Tang YT, Liu C;
 PI Drmanac RT, Wang M, Chen L, Yang Y;
 XX WPI: 2001-465705/50.
 DR
 XX
 PT Novel metallocarboxy peptidase-like polypeptide is useful in treating
 PT neurodegenerative, immune, etc. disorders -
 XX
 XX
 PS Example 3; Fig 1; 146pp; English.
 XX
 CC The invention relates to an isolated polypeptide comprising an amino
 CC acid sequence of 374, 41, 14, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
 CC 354, 165, 374, 354, or 41 amino acids, fully defined in the
 CC specification, or its translated protein coding portion. Mature protein
 CC coding portion, extracellular portion or active domain. The polypeptide
 CC is a human secreted metallocarboxypeptidase-like polypeptide. It
 CC is useful for treating mammalian diseases including:
 CC neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
 CC osteoporosis, osteoarthritis, disorders linked to reduced tissue
 CC growth, Huntington's disease, amyotrophic lateral sclerosis, stroke,
 CC reperfusion, wound healing, infectious diseases, immune and autoimmune
 CC disorders such as severe combined immunodeficiency, rheumatoid
 CC arthritis, Guillain-Barre syndrome, insulin dependent diabetes
 CC mellitus, graft-versus-host disease and allergy, cancer, and
 CC haemophilia. The polypeptide is also useful in medical imaging, as an
 CC immunogen to generate antibodies, as a nutritional source or
 CC supplement, and as a contraceptive. The present sequence is
 CC a carboxypeptidase homologue protein of Bothops jararaca.
 CC
 XX
 SQ Sequence 324 AA;
 XX
 Query Match 32.9%; Score 629.5; DB 22; Length 324;
 Best Local Similarity 41.7%; Pred. No. 6.2e-56;
 Matches 125; Conservative 53; Mismatches 79; Indels 43; Gaps 5;
 XX
 QY 84 LNVSGIPCSVLADVEDLIQOQISNDTSPRASASYEYQHSNLETYSWIEFTERRHPD 143
 DB 2 lqsgslngyellldnlgavldqldha---rtagynyekynswekidwadidanemsl 58
 QY 144 LTKIHIGSSPEKRYPLVAVSGKEQTAKNAIWDGCIHAREWISPAFLMFT----- 195
 DB 59 vsrlqigtltfegrmpplkv-gkpgvnxkalfidcgfharewelspafcgwfvreaavrtly 117
 QY 196 -----GNNRMWRKRNSFYANNHCIGTDLSNFFYSK 226
 DB 118 ketlmtqllnkldfyilpvlndgyvyswksrmwrtkrsvnaegstcigtprnrf-daa 176
 QY 227 WCEEGASSSSCSEYTCGLYPESEPEVKAVASFLRRINQIKAYISMHSYSQHIYFPYSY 286
 DB 177 wcvsgasrnpesecygcgkpeseketkaladfirnrslsqaylthhsyqgmmllypsyt 236

QY 287 RSKSKHSELVLVASEAVRAIDKSKNRTYTHGSEFTLVLAPGGDDWYDLGKYSFT 346
 Db 237 ydltsnmknklnslakelrei-kvligtelytgpgaaltlypaagsddwagdgqlykyaft 295

RESULT 14

AAW41414
 ID AAW41414 standard; Protein; 716 AA.

AC AAW41414;

DT 02-JUN-1998 (first entry)

DE PreproHCPB-linker-humanised 806.077 VH.

KW Anti-CEA antibody; carcinoembryonic antigen; 806.077 Ab; cancer therapy;
 cancer diagnosis; complementarity determining region.

OS Synthetic.

PN W09742329-A1.

PD 13-NOV-1997.

PF 29-APR-1997; 97MO-GB01165.

PR 14-FEB-1997; 97GB-0003103.

PR 04-MAY-1996; 96GB-0009405.

PA (ZENE) ZENECA LTD.

PI Copley CG, Edge MD, Emery SC;

DR WPI: 1997-558987/51.

DR N-PSDB; AAV17331.

PT Anti-carcinoembryonic antigen antibody 806.077 Ab - used for

diagnosis and therapy of cancer

Example 75; Page 184-187; 208pp; English.

This sequence is preproHCPB-linker-humanised 807.077 VH, which is an example of the antibody of the invention. The antibody is an anti-CEA (carcinoembryonic antigen) antibody (preferably 806.077 Ab). Host cells or transgenic organisms transformed with DNA encoding the antibody, are used to make the antibody or conjugate. The conjugate is used in a medication suitable for intravenous administration. The conjugate can be used for cancer therapy, selectively killing tumour cells. The antibody can be used for in vivo or in vitro diagnosis of cancer.

Sequence 716 AA;

Query Match 32.6%; Score 623; DB 18; Length 716;

Best Local Similarity 34.6%; Pred. No. 1,1e-54;

Matches 132; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

QY 3 LCSLAVLYPIVLFCEQHFA-FQSGVLAALPRTSRQOVQLONTTTEIYIMQPVADL 61

Db 1 mhallivltvalasahghegekefvrnvdenhnlirelastcqidfwkpsvtq 60

QY 62 IVKKQVFEFVNASDVNDVKAHLNVSIGIPSVLADVEDLIDQOISNDTPSRASAYE 121

Db 61 lkphtvfrkaedvtvenvlnkqnelgykvlslntnveagdsrv--ratghsye 117

QY 122 QYHSLNETYSWIEFTTERHPDMLTKIHGSEFEKYPYLVLKVSQKEQYAKNAIWDGII 181

Db 118 kynkwtelavtqvatentpalrsvlgttfegrellyllkv-gkaqgnkpalfndcgfh 176

QY 182 AREWISPAFLCMFI-----GH-----NDMNRKN 204

Db 177 arewispafcgwfvreavtlygrtqvelldkldfyvlpvlnldgylytwcksrftwkt 236

QY 205 RSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSEFTYCGLYPSEPEVKAASFRLRNIN 264
 Db 237 rslhtgssclgtprntnf-dagwcelgaarncpdcelycgpaaeseketaiaadfnkis 295

QY 265 QIKATYSMHSYSOHIVPPYSYTRSKSDHELSLVASEAVRAIDKSKNRTYTHGSEFT 324

Db 296 slkayiltlnysgqmilypysayklgenaelnalakavkel-aahgtkytlypggatt 354

QY 325 LYLAPGGDDWYDLGKYSFT 346

Db 355 lypsagtskdwagdgqlyrft 376

RESULT 15

AAW00602
 ID AAW00602 standard; Protein; 307 AA.

AC AAW00602;

DT 09-NOV-1996 (first entry)

DE Rat carboxypeptidase B mature enzyme.

KW Carboxypeptidase-B; pro-enzyme; protease; insulin.

OS Rattus sp.

PN W09623064-A1.

PD 01-AUG-1996.

PF 25-JAN-1996; 96MO-US00995.

PR 25-JAN-1995; 95US-0378233.

PA (BIOT-) BIOTECHNOLOGY GEN CORP.

PI Fulga N, Gorecki M, Hartman J, Mendelovitch S;

DR WPI: 1996-362688/36.

DR N-PSDB; AAT35760.

PT Purified active recombinant carboxypeptidase B prodn. - by expressing DNA encoding the pro-enzyme, folding and enzymatic cleavage to give active form, used e.g. for insulin prodn.

Example 1; Page 37-38; 49pp; English.

The amino acid sequence of the rat mature carboxypeptidase-B (CPB) (AAW00602) was deduced from a cDNA sequence (see also AAT35760) obtd. by PCR amplification of Sprague-Dawley rat pancreas cDNA. The sequence of the activation peptide (AAW00601) of pro-CPB has also been deduced. Pro-CPB (pref. with a C290S mutation) can be expressed in host cells, e.g. Escherichia coli, and subsequently recovered, refolded and cleaved with trypsin to yield the active enzyme. CPB produced this way is cheaper than porcine pancreatic enzyme, and is free of other proteases.

Sequence 307 AA;

Query Match 31.6%; Score 604; DB 17; Length 307;

Best Local Similarity 44.6%; Pred. No. 2,4e-53;

Matches 120; Conservative 43; Mismatches 66; Indels 40; Gaps 5;

QY 115 ASASYEIOYHSLNETYSWIEFTTERHPDMLTKIHGSEFEKYPYLVLKVSQKEQYAKNAI 174

Db 1 asgnsytkynnmetleawlgvatdnpdlvqvsygttfggrmyviki-gktrpnkpal 59

QY 175 WIDGIAHREWISPAFLCMFI-----GH----- 197

Db 60 fidegfharewispafcgwfvreavtlygrtqvelldkldfyvlpvlnldgylytwck 119

QY 198 NRMWRKRSPFYANNHCIGFDLNSNFVSKHWCCEGASSSCSEYCGLYPSEPEVKAVAS 257
Db 120 drrmrkLrslmagssclgvprnrf-nagwcevgasrpscsetycgpapeseketka1ad 178
QY 258 FLRRNINQIKAYISMHSYSOHIVPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYT 317
Db 179 flrrn1stlkayltlthysqgmlypyeydyklpenyeelnalvkaakel-atlhgtkyt 237
QY 318 HGHGSETLYLAPGGGDWYIDLGIKYSFT 346
Db 238 yppgatllypaagsddwsydgikysft 266

Search completed: September 18, 2002, 16:45:11
Job time: 378 sec

This Page Blank (uspto)

PT haemophilia A
 XX
 PS Disclosure: Figure 4; 40pp; English.
 CC
 CC An antibody which specifically binds human plasma carboxypeptidase B
 CC (hPCPB) and does not cross react with other carboxypeptidases is
 CC useful for the detection of hPCPB in vitro. The antibody is also
 CC used for purifying hPCPB from a sample. Purification comprises
 CC passing a sample thought to contain hPCPB over either a column to
 CC which antibody has been bound, or a plasmaogen affinity column,
 CC eluting the column and then recovering the fraction containing the
 CC hPCPB.
 XX
 SQ Sequence 423 AA;
 alignment_scores:
 Quality: 1885.00 Length: 423
 Ratio: 5.122 Gaps: 2
 Percent Similarity: 86.998 Percent Identity: 86.525
 alignment_block:
 US-09-980-881-1 x AAR90293 ..
 Align seg 1/1 to: AAR90293 from: 1 to: 423
 18 ATGAAGCTTTCAGAGCTTGCAGTCTGTACCATTTGCTCTCTCTGTA 67
 1 MetylsleucysserleuAlaValleuValProilleuValleuPheCysgl 17
 68 GCAGCAGTCTTCGCGTTCAGAGAGTGGCCAAAGTTTACGCTCTCTCTTA 117
 17 uGlnHisValAlpheAlaPheGlnSerGlyGlnValleuAlaAlaLeuProA 34
 118 GAACCTTAGGCAAGTCAAGTTTACAGAACTTACTACAAATATATAG 167
 34 rGlnSerAlaGlnValGlnValleuGlnAsnLeuThrThrTyrGln 50
 168 ATGTGTTCTGCGCAGCGGTAAACAGTACCTTATTTGAGAAAAACA 217
 51 lleValleuTrpGlnProvalThrAlaAspLeuIleVallyslysGln 67
 218 AGTGCATTTTGTAAATGATCTGATGTGAGAAATGTGAAGCCCAT 267
 67 nValHisPhePheValAsnAlaSerAspValAlaAsnValAlaHisL 84
 268 TAAATGTAGGAGGAAATTCATGAGTGTCTGTGGCAGAGCTGGAAGAT 317
 84 euAsnValSerGlyIleProCysSerValleuLeuAlaAspValGlnAsp 100
 318 CTATTCACAGCAGATTTCCACAGACAGTACGCCCGCCGAGCCTCCGC 367
 101 leuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaGlnAsnAl 117
 368 ATCTACTATGACAGATGATCAGTCACTAAATGAATCATTTCTGGATAG 417
 117 aserTyrTyrGlnGlnIleTyrHisSerLeuAsnGlnIleTyrSerTrpIleG 134
 418 AATTATTAAGTACAGAGCATCCTGATATGTTTACAAAAATACATTTGA 467
 134 luePheIleThrGlnAlaGlnHisProAspPheLeuThrLysIleHisIle 150
 468 TCTCATTTGAGAGTACCCACTATGTTTAAAGTTTCTGGAAAAAGA 517
 151 SerSerPheGlnLysTyrProLeuTyrValleuLysValSerGlyLysGln 167
 518 ACAAAACGCGCAAAATGCCATATGATGATGAGTGGAAATCCATGCCAGAG 567
 167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyLleHisAlaArg 184
 568 AATGATCTCTCCGCTTCTGCTTGTGTCATAGGCCAT..... 608
 184 lntTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

608 608
 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspH 217
 608 608
 217 eTyrValMetProValValAsnValAspGlyTyrAspTyrSerTrpLysL 234
 609 ..AATCGAATGTGGAGAAAGAACCGTTTCTTCTATGCAACAAATCATGCG 656
 234 ysAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCys 250
 657 ATCGAAGACAGACTGGAATAGCAACTTGTCTCCAAACACTGGGTGAGGA 706
 251 IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGlnGln 267
 707 AGCGCATCCAGTTCCTCATGCTGCGAAACCTACCTGCTGAGCTTATCTGTG 756
 267 uGlyAlaSerSerSerSerCysSerGlnThrTyrCysGlyLeuTyrProG 284
 757 AGTCAGAACAGAAAGTGAAGCAGTGGCTAGTTCTTGAGAAAGAAATATC 806
 284 lueSerGlnProGlnValLysAlaValAlaSerPheLeuArgArgAsnIle 300
 807 AACGATTTAAAGCATACATGACATGATCATATCTCCAGCATATAGT 856
 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 317
 857 GTTTCATATTCCTATACACGAGTAAAGCAAAAGCAACATGAGAACTGT 906
 317 lPheTrpTyrSerTyrThrArgSerLysSerLysAspHisGlnGlnLeuS 334
 907 CTCTAGTACGACAGTACAGACAGTGTCTATTTGACAAACTAGTAAAT 956
 334 eTleuValAlaSerGlnAlaValAlaArgAlaIleGlnLysThrSerLysAsn 350
 957 ACCAGGTATACACATGCCATGCTCAGAAACCTTATACCTAGCTCTGG 1006
 351 ThrArgTyrThrHisGlyHisGlySerGlnThrLeuTyrLeuAlaProGln 367
 1007 AGCTGGGCGACATTTGATCTATGATTTGGCATCAAAATTTCTGTT..... 1051
 367 yGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThrI 384
 1052TAC 1054
 384 leGlnLeuAlaArgAspThrGlyThrTyrGlyPheLeuLeuProGlnArgTyr 400
 1055 ATCAAAACCCACCTGTAGAGAAAGCTTTGCCGCTGTCTTAAATAGCTTG 1104
 401 IleLysProThrCysArgGlnAlaPheAlaValSerLysIleAlaIle 417
 1105 GCATGTCATTTAGCAATGTT 1123
 417 PHisValIleAlaArgAsnVal 423
 seq_name: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1197.DAT:AAW14733
 seq_documentation_block:
 ID AAW14733 standard; Protein: 423 AA.
 XX
 AC AAW14733;
 XX
 DT 08-MAY-1997 (first entry)
 XX
 DE Human plasma carboxypeptidase B.
 XX
 KW Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
 KW plasma; plasmaogen.
 XX
 OS Homo sapiens.
 XX

Key	Location/Qualifiers
Peptide	1..22
Protein	/note= "Signal peptide" 23..423
Cleavage-site	/note= "Mature PCPB" 114
Active-site	/note= "Potential clip site for activation of PCPB as a carboxypeptidase" 181
Binding-site	/note= "Expected to be involved in catalytic activity" 183
Active-site	/note= "Expected to be involved in substrate binding" 184
Binding-site	/note= "Expected to be involved in catalytic activity" 256
Binding-site	/note= "Expected to be involved in substrate binding" 257
Active-site	/note= "Expected to be involved in substrate binding" 310
Binding-site	/note= "Expected to be involved in catalytic activity" 312
Active-site	/note= "Expected to be involved in substrate binding" 369
Binding-site	/note= "Expected to be involved in catalytic activity" 370
Blinding-site	/note= "Expected to be involved in substrate binding, thought to determine specificity of PCPB as a carboxypeptidase B" 371
Binding-site	/note= "Expected to be involved in substrate binding" 394
US5593674-A.	/note= "Expected to be involved in substrate binding" /note=
14-JAN-1997.	
01-FEB-1991;	91US-0649591.
01-FEB-1991;	91US-0649591.
14-OCT-1992;	92US-0959944.
15-DEC-1993;	93US-0167727.
19-JUL-1994;	94US-0275540.
27-APR-1995;	95US-0430787.
(GETH) GENENTECH INC.	
Drayna DT, Eaton DL;	
WPI: 1997-099413/09.	
N-PSDB: NAT62846.	
Using human plasma carboxy:peptidase B in blood coagulation - is functionally related to carboxy:peptidase A and pancreas carboxy:peptidase B	
Example 2; Column 37-42; 39pp; English.	
This sequence represents human plasma carboxypeptidase B (PCPB) which has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD. PCPB may be used therapeutically in haemostatic regulation. PCPB is purified from human plasma or by transformed cell culture by extraction using plasminogen bound to a solid phase.	
Sequence	423 AA;

Alignment_scores:

Quality: 1885.00 Length: 423

Ratio: 5.122 Gaps: 2

Percent Similarity: 86.998 Percent Identity: 86.525

Alignment_block:

US-09-980-881-1 x AAW14733 ..
Align seg 1/1 to: AAW14733 From: 1 to: 423

18	ATGAAGCTTTGCAAGCCTTGACAGCTCTGTGACCAATGTTGCTCTGTGGA	67
1	MethylScusserleuAlaValLeuValProIleValLeuPheCysGcl	17
68	GCACAGATGCTTGCGCGTCCAGAGTGGCCAGTGTACGTGCTTCTCTA	117
17	uGlnHisValPheAlaPheIleSerGlyIleValLeuAlaIleuProA	34
118	GAACCTTAGCAAGTTCAAGTCTACAAATCTTAACAAACATATGAG	167
34	rglnSerAlaArgIleValGlnValLeuGlnAsnLeuThrThrThyTyrGln	50
168	ATTGTTCTCTGGACGCGGATGACAGCTGACCTTATGTGAAGAAAAACA	217
51	IleValLeuThrProIleProValThrAlaSerLeuIleValIlylCysGcl	67
218	AGTCCATTTTTTTGTAAATGATCTGATGTGCAACATGTGAAGCCCAT	267
67	nValHisPhePheValAlaAsnIaSerAspValAlaSerValIlylAlaHisL	84
268	TAAATGTGACGGCAATTCATGACAGTGTGCTGCGACAGCTGGAGAT	317
84	eukAnValSerGlyIleProCysSerValLeuLeuAlaIaSerValGlnAsp	100
318	CTTATTCACACAGAGATTTCACACGACACAGTCAAGCCCCGACGCTCGC	367
101	LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaIaSerAl	117
368	ATGGTACATGACAGCTATACATCACTCAATGAATCTATCTTGATAG	417
117	aseTyrTyrGlnIleGlnIlylHisSerLeuAsnGlyIleTyrSerTrpIleG	134
418	AATTTATACAGAGAGCATCTGATATGCTTACAAATGCACATTTGGA	467
134	IurPheIleThrGlnIurGlnHisProAspMetLeuThrIlylIleHisIleGly	150
468	TTCCTCATTTGAGAAGTACCACCTGTATGTGTTTAAAGTTCTGGAAAA	517
151	SerSerPheGlyIlylSerTyrProLeuTyrValLeuIlylValSerGlyIlylGln	167
518	ACAACAAGCCAAAATAGCCATATGAGTATGACGTGGGAATCCATCCGAG	567
167	uGlnIleThrAlaIlylAsnAlaIleIleTrpIleAspCysGlyIleHisAlaTrg	184
568	AATGATCTCTCTGCTTTCTGCTGTGTGATCATTAGCCAT	608
184	IurTrpIleSerProAlaPheCysLeuThrPheIleGlnHisIleThrGln	200
608	608
201	PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuAlaTrgLeuValAsp	217
608	608
217	eTyrValMetProValIleAlaSerValAlaSerGlyTyrAspTrpSerTrpIlyl	234
609	.. AATGAATGTGAGAAAGAACCGTCTTTCTATGCAACAATCATTTGC	656
234	yAsnAlaTrgMetTrpAlaTrgIlylAsnAlaTrgSerPheTyrAlaAsnAlaHisCys	250
657	ATGGGAACGACCTGAATATGACAACTTTGTCTCCAAACAGTGGTGTAGGA	706
251	IleGlyThrAspLeuAsnAlaTrgAsnPheAlaSerIlylHisIleTrpCysGlnG	267
707	AGGTGCATCCAGTTCTCATGCTCGGAACCTACTGTGCACTTATCTTG	756
267	uGlyAlaSerSerSerSerCysSerGlnThrThyCysGlyIleuTyrProG	284
257	AGTCAGACACCAAGTGAAGCGAGTGGCTAGTTTCTTGAGAGAATATTC	806


```
|||||
284 lusergluprogluvalalalaserpheuleuargashile 300
807 AACCATTAAGCATATCATCATCTATCTCCAGCATATAGT 856
301 AsnGlnIleuysAlatylIleSerMetHisSerlySerGlnHisIleVa 317
857 GTTTCATATTCCTATACAGAAATAAGCAAGACCATGAGGACTGT 906
317 lPhePtyrSerlyThrArSerlySerlySAspHisGluIleuS 334
907 CTCTAGAGCAGAGAGAGTCTGCTATTTGCAAAACTAGCAAAAT 956
334 erleuvalAlaserGluAlaValaAlaIleGluIlyThrSerLyasn 350
957 ACCAGATATACATGCGCATGCGTCAGAAACCTTATACCTAGCTCTGG 1006
351 ThrArglyThrHisGlyHisGlySerGluThrleuTyrlleAlaProG 367
1007 AGGTGGGAGCATTTGATCTATGATTTGGGCATCAAAATATTCGTT 1051
367 YGlyGlyAspAspTrpIleTyraSpleuGlyIleLySerPheThrI 384
1052 .....TAC 1054
384 legIuleuArgAspThrGlyThrTyrlPheleuProGluArgTyr 400
1055 ATCAAAACCCACCTGTAGAGAGCTTTGGCGCTCTCTAAATAGCTTG 1104
401 lIlelySerThrCysArgGluAlaPheAlaIleValSerlyIleAlaTr 417
1105 GCATGTCATTAGGAATGTT 1123
417 PHISvalIleArgAsnVal 423
seq_name: /stds1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:AAW92270
seq_documentation_block:
ID AAW92270 standard; Protein; 423 AA.
AC
AAW92270;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KM polymorphism detection; thrombotic disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT Protein /note="signal peptide"
FT /note="mature PCPBthr147"
XX
PN WO9855645-A1.
XX
PD 10-DEC-1998.
XX
PF 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD ) SCHERING AG.
XX
PI Morser MJ, Nagashima M:
XX
XX WPI; 1999-045800/04.
XX DR N-PSDB; AAV74302.
XX
PT Detecting new polymorphism of human plasma carboxypeptidase B -
```

```
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
PS Disclosure; Page 25-26; 35pp; English.
XX
CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
XX
SQ Sequence 423 AA;
XX
alignment_scores:
Quality: 1885.00 Length: 423
Ratio: 5.122 Gaps: 2
Percent Similarity: 86.998 Percent Identity: 86.525
alignment_block:
US-09-980-881-1 x AAW92270 ..
Align seq 1/1 to: AAW92270 from: 1 to: 423
18 ATGAAGCTTTGACAGCTTGAGTCCTTACCAATGTTCTCTTCTGGA 67
1 MetIlySerIleuysSerleuAlaValleuValProIleValleuPheCysG 17
68 GCAGCATGCTCTTCGCTGCAGAGTGGCCAAATTCTAGCTCTCTTCCTA 117
17 uGlnHisValAlaPheAlaPheGlnSerGlyGlnValleuAlaIleuProA 34
118 GAACCTCTAGGCAAGTTAAAGTTCTACAGAACTTACTACAACTATGAG 167
34 rGlnHisSerArgGlnValGlnValleuGlnAsnLeuThrThrTyrlGlu 50
168 ATTGTTCTCTGGCAGCCGTAACAGCTGACCTTATTTGGAAGAAAACA 217
51 lIleValleuTrpGlnProvalThrAlaAspleuIleValIySlySyl 67
218 AGTCATTTTGTGTAATGATCTGATGTCAGCAATGTGAAGCCCAT 267
67 nValHisPhePheValAsnAlaSerAspValAspAsnValIySlyAlaHisL 84
268 TAAATGTGAGCGGAATTCATGACAGTGTCTGTGGCAGAGTGGAAAT 317
84 euAsnValSerGlyIleProCysSerValleuLeuAlaAspValGluAsp 100
318 CTATATCAACAGAGATTTCCAAGCAGACAGTCAAGCCCGAGCCTCGC 367
101 leuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaIleSerAl 117
368 ATGCTACTATGAGACAGTATCACTCACTAATGAAATCTATCTTGATG 417
117 aSerTyrlTyrlGluGlnIyHisSerleuAsnGluIleTyrlSerTrpIleG 134
418 AATTATACAGAGAGCATCTGATATGCTTACAAAATCCACATATGGA 467
134 lIupheIleThrGluArgHisProAspMetLeuThrIyHisIleHisIleI 150
468 TCTCATATTGAGAGTACCACTCATGTTTAAAGTTCTGGAAGAAAGA 517
151 SerSerPheGluIyTyrlProleuTyrlValleuIySylSerGlyIySyl 167
```

```

518 ACAAAGGCGGAAAGAAATGCAATGAGATGAGTGGATGCAATGCGGAGAG 567
167 uGnTThAlaLysAsnAlaIleTrrPLeaSRySGuYIleNiSaIaLaTgG 184
568 AATGAGTCTCTCCGCTTTCTGCTTGGCTTCAATAGGCAAT..... 608
184 LuTrrPLeaSRProLaRheScyLeuEutrrPheIleGlnIleThrGln 200
608 ..... 608
201 PheTgYGLYleIleGIGlnTYThraSnlLeuEaLgLeuValaSprh 217
608 ..... 608
217 eTYThValMeSRProValaLAsnValaSPRGLYThraSPryTSeTrrPLeaL 234
609 . . AATGCAATGTGGAGAAAGCAACGGCTTTCTTATGCGCAACATCAATGCG 656
234 yAsnAlaGMeTrrPrrGLyAsnAlaGSeTrrPheTYThraSnlAsnIlaScYs 250
657 ATCGGAACAGACSTGAATATGCAACTTTGTCTCCAAACSTGCTGTGAGAA 706
251 IlGGLYThraSPryLeuSnlAaGAsnPhraIlaSerYlNiSaTrrcyaGLuG 267
707 AGGGAGTCCAGTTCCTCAATGCTCGGAAACSTACTGTGGACTTTATCGCG 756
267 uGyAlaIaSeTSeTSeTSeTSeTSeTSeTSeTSeTSeTSeTSeTSeTSeT 284
757 AGTCAGAACAGCAAGTGAAGAGGCAAGTGGCAAGTTCTTGGAGAAATATGC 806
284 LuSeTrrGLProGLYValaLValaValaIaSeTrrPheLyuLgrraGsnIle 300
807 AACCAAGATTAAACATATCATCATGACATGCAATTCATCTCCACATATAGT 856
301 AsnGlnIleLysAlaTYThraSPryLeaSeTrrMetIlaSeTrrSerGlnIleVa 317
857 GTTTCATATTCCTATACACAGATGAAGATGAAGCAAGCAAGCATGAGAACTG 906
317 LrrPrrPryTSeTrrTYThraTgSeTrrLysSeTrrLysAsnIlaGlnGluLeuS 334
907 CTCTAGTACCACTGAAGACAGTTCGCGCTATTGACAAACATAGTAATAAT 956
334 eTrrLyuValaIaSeTrrGLYValaLValaLValaIleGluYSThSeTrrLysAsn 350
957 ACCAGATATACATAGGCGCATGGCTCCAGAAACSTTATACSTAGTCTCTGG 1006
351 ThraTgTYThraISeGlnIleGlnIleGlnSeTrrGlnThLeuTYLeuAlaProGL 367
1007 AGGTGGGGAGCAATGGATGATGATGATTTGGGCAATCAATATTGCTT.... 1051
367 yGlnGLYAsnAsnTrrPrrLeuTYAsnProGLYLeuYSYSeTrrPheThrI 384
1052 ..... TAC 1054
384 leGluLeuLgrraSPrrTrGLYThrTYGlnYrPheLeuLeuProGLYArTYG 400
1055 ATCAAAACCCACSTGTAGAGAAACSTTTTGGCGCTGTCTCTAAATAGCTTG 1104
401 IlLysProThrCyArGLYValaLPhraIlaValaIaSeTrrLysIleAlaTr 417
1105 GCATGTCATATGAAATGTT 1123
417 pHISValIleAlaGAsnVal 423

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: AAR36273

seq_documentation_block:

ID	AA
AAR36273	standard; Protein; 423 AA.

AC AAR36273;

DT 24-AUG-1993 (first entry)

XX

DE	Human plasma carboxypeptidase B.		
xx			
KW	PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.		
OS			
XX	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	Peptide	1..22	
FT		/label= signal_peptide	
FT	Cleavage-site	114	
FT		/note= "putative site of cleavage by trypsin to activate PCPB as a carboxypeptidase"	
FT	Active-site	181	
FT		/note= "expected to be involved in catalytic activity"	
FT	Binding-site	183	
FT		/note= "expected to be involved in substrate binding"	
FT	Active-site	184	
FT		/note= "expected to be involved in catalytic activity"	
FT	Binding-site	256..257	
FT		/note= "expected to be involved in substrate binding"	
FT	Active-site	310	
FT		/note= "expected to be involved in catalytic activity"	
FT	Binding-site	312	
FT		/note= "expected to be involved in substrate binding"	
FT	Active-site	369	
FT		/note= "expected to be involved in catalytic activity"	
FT	Binding-site	370..371	
FT		/note= "expected to be involved in substrate binding; Asp 370 is the residue which determines specificity of the PCPB as a carboxypeptidase B"	
FT	Binding-site	394	
FT		/note= "expected to be involved in substrate binding"	
XX			
PN	US5206161-A.		
XX			
PF	27-APR-1993.		
XX			
PF	01-FEB-1991; 91US-0649591.		
XX			
PR	01-FEB-1991; 91US-0649591.		
XX			
PA	(GENE) GENENTECH INC.		
XX			
P1	Drayna DT, Baton DL;		
XX			
DR	WPI; 1993-151724/18.		
DR	N-PSDB; AA041001.		
XX			
PT	New human plasma carboxypeptidase B - used as haemostatic		
XX	regulator for clotting blood, partic. for treating blood clotting		
PT	disorders, e.g. haemophilia		
PS			
XX	Claim 2; Fig 4; 40pp; English.		
CC			
CC	This amino acid sequence was deduced from a clone isolated from a		
CC	human liver cDNA library. There is about 40% sequence identity		
CC	between the deduced preproPCPB amino acid sequence and pro-human		
CC	mast cell carboxypeptidase A and between preproPCPB and prepro-rat		
CC	carboxypeptidase B. Human PCPB has the same substrate binding site		
CC	as, and shares 6 cysteine residues which form 3 intramolecular		
CC	disulphide bonds with bovine and rat carboxypeptidase B. The		
CC	presence of the same amino acid (asp at position 348 of the mature		
CC	protein) at the region in carboxypeptidases that determines		
CC	substrate specificity, suggests that PCPB represents a plasma-		

CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC LPA of plasminogen to plasmin in the presence of fibrinogen.

XX
SQ Sequence 423 AA;

alignment_scores:
Quality: 1880.00 Length: 423
Ratio: 5.123 Gaps: 2
Percent Similarity: 86.761 Percent Identity: 86.288

alignment_block:
US-09-980-881-1 x AAR36273 ..

Align seg 1/1 to: AAR36273 from: 1 to: 423

```

18 ATGAGGCTTTCAGCCTTCAGCTTCCTGACCATTCCTTCCTTCCTGCA 67
   |||||||
1  MetLysLeuCySerLeuAlaValLeuValProLeuValLeuPheCysG1 17
68 GCACCATGCTTCGCGTTCAGAGTGGCCAAATTCTAGCTGCTTCCTTA 117
   |||||||
17  uGlnHisValPheAlaPheGlnSerGlnValLeuAlaIleLeuProA 34
118 GAACCTTGAGCAAGTTCAGTTCACAGAACTTCTACACATPAGAG 167
   |||||||
34  rGTHSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGln 50
168 ATTGTCCTGCGCAGCCGGTACAGCTGACCTTATGTGAGAAAGAAACA 217
   |||||||
51  IleValLeuTrpGlnProValThrAlaAspLeuValLysLysG1 67
218 AGTCCATTTTTCGTAATGCATCTGATGTCAGACATGTGAAGCCCAT 267
   |||||||
67  nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
268 TAAATGAGGCGGAATTCATGCGAGTGTCTGTCGGCAGACGTGGAGAT 317
   |||||||
84  euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGlnAsp 100
318 CTATTTCACAGCAGATTTCACAGCAGACAGTCCAGCCCGAGCCTCGC 367
   |||||||
101  LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaGlnSerAl 117
368 ATCTACTATGACAGATATCACTCACTAAATGAATCTATTCTTGATAG 417
   |||||||
117  aSerTyrTyrGlnGlnIleThrHisSerLeuAsnGlnIleTyrSerTrpIleG 134
418 AATTATTAACAGAGCAGCATCTGATATGCTTACAAAATCCACATTGGA 467
   |||||||
134  IurPheIleThrGlnArgHisProAspMetLeuThrLysIleHisIleGly 150
468 TCCCATTTTGAGAGTACCCCACTATGTTTAAAGTTTCTGGAAGA 517
   |||||||
151  SerSerPheGlnLysTyrProLeuTyrValLeuLysValSerGlnLysG1 167
518 ACAACAGCCAAATATGCATATGATGACTGTGGAATCCATGCCAGAG 567
   |||||||
167  uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyLleHisAlaArg 184
568 AATGAGATCTCTCGCTTCCTGCTGCTGGTTCATAGGCCAT..... 608
   |||||||
184  IurPheSerProAlaPheCysLeuTrpPheIleGlnHisIleThrGln 200
608 ..... 608
201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuAlaGlnLeuValAspPh 217
608 ..... 608
217 eTyrValMetProValValAsnValAspGlyTyrAspTyrSerTrpLysL 234
609 ..AATCGAATGTGAGAAAGAACCGTCTTCTATGCGAACAATCATTTGC 656

```

```

234 yAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCys 250
657 ATCGACAGACCTGATAGCAACTTGTTCACCAACACTGGTGAGAGA 706
   |||||||
251  IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGlnG1 267
707 AGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGAGATTATCCGTG 756
   |||||||
267  uGlyAlaCysSerSerSerCysSerGlnThrTyrCysGlyLeuTyrProg 284
757 AGTCAGAACCCAGAGTGAAGCAGTGGCTAGTTTCTTGAGAAAGAAATTC 806
   |||||||
284  IuSerGlnProGlnValLysAlaValAlaSerPheLeuArgArgAsnIle 300
807 AACCATTTAAAGCATACATGACATGACATTCATCTCCAGCATATAGT 856
   |||||||
301  AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVa 317
857 GTTTCATATTCCTTACACGAAAGTMAAGCAAGCAATGAGAACTGT 906
   |||||||
317  IpheTrpTyrSerTyrThrArgSerLysSerLysAspHisGlnGlnLeuS 334
907 CTCTAGTACCCAGTGAACAGTTCGCTGCTATTGACAAACTAGTAAAT 956
   |||||||
334  eTyrLeuValAlaSerGlnAlaValArgAlaIleGlnLysThrSerLysAsn 350
957 ACCAGTATACACATGCGCAGTGGCTCAGAAACCTTATACATGCTCCG 1006
   |||||||
351  ThrArgTyrThrHisGlyHisGlySerGlnThrLeuTyrLeuAlaProG1 367
1007 AGGTGGGAGCAGATTGATCTATGATTGGGCAATCAATATTCGTT..... 1051
   |||||||
367  yGlyLysAspAspTrpIleTyrAspLeuGlyLleLysTyrSerPheTrpI 384
1052 .....TAC 1054
384  IeGlnLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGlnArgTyr 400
1055 ATCAAAACCACCTGTAGAGAGCTTTGCGGCTGCTCTAAATAGCTTG 1104
   |||||||
401  IleLysProThrCysArgGlnAlaPheAlaValAlaValSerLysIleAlaTr 417
1105 GCATGTCATTAGGAATGTT 1123
   |||||||
417  PHisValIleArgAsnVal 423
seq_name: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT: AAB11458
seq_documentation_block:
ID AAB11458 standard; Protein: 338 AA.
XX AAB11458;
AC
XX
DE Human brain carboxypeptidase B protein SEQ ID NO 3.
XX
XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
XX cerebroprotective; antiAlzheimers; neurotropic; neuroprotective;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma.
XX
XX Homo sapiens.
XX
XX WO2000066717-A1.
XX
XX 09-NOV-2000.
XX
XX 01-MAY-2000; 2000MO-JP02878.
XX
XX 30-APR-1999; 99UP-0125169.
XX
PR

```

XX (MATS/) MATSUMOTO A.
 XX
 XX Matsumoto A;
 XX
 XX WPI; 2000-687534/67.
 XX
 PT Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX
 PS Disclosure: Page 71-73; 84pp; Japanese.
 XX
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity, and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 XX
 SQ Sequence 338 AA;

alignment_scores:
 Quality: 1785.00 Length: 338
 Ratio: 5.297 Gaps: 0
 Percent Similarity: 99.704 Percent Identity: 99.112

alignment_block:

US-09-980-881-1 x AAB11458 ..

Align seg 1/1 to: AAB11458 from: 1 to: 338

84 TTCCAGAGTGGCCAAAGTTTCAGGCTTCCTAGAACCTCTAGGCAAGT 133
 1 PheGlnSerIcylGlnValLeuAlaAlaLeuProArgThrSerArgGlnVal 17
 134 TCAAGTTCTACAGAACTTCTACACATATGAGATTGTTCTCTGCGAGC 183
 17 IGIValLeuGlnAsnLeuThrThrTrpGluIleValLeuLeuTrpGlnP 34
 184 CGGTACAGCTGACCTTATTTGTGAAGAAAAACAGTCCATTTTGTGA 233
 34 rovalThrAlaAspLeuIleValIleValIleValIleValIleVal 50
 234 AATGATCTGATGTGCGAATGTGAAGCCCATTTAAATGTGACGGAAT 283
 51 AsnAlaSerIleAspValAspAsnValIleValIleValIleValIleVal 67
 284 TCCATCAGTGTCTGTGCGAGAGCGTGAAGATCTTATTCACAGCAGA 333
 67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
 334 TTTCGAAGCAGACAGTCAGCCCGGAGCCCTCCGCACTGTAATGTAACAG 383
 84 IeSerAsnAspTrpValSerProArgAlaSerAlaSerIleValIleVal 100
 384 TATCACTCACTAAATGAATCTATTTGGATGAGATTATTAACAGAGAG 433
 101 TyrHisSerLeuAsnGluIleValIleValIleValIleValIleVal 117
 434 GCATCTGTATGCTTACAAAAATCCACATGTGATCTCATTTGAGAGAGT 483
 117 gHisProAspMetLeuThrIleHisIleGlySerSerPheGluIleVal 134
 484 ACCCACTCATATGTTTAAAGTTCTCGAAGAAAGAACAAACAGCAAAAT 533
 134 yPProLeuIleValLeuIleValSerGlyIleValIleValIleValIleVal 150
 534 GCCATATGATGTGACTGTGAAATCCATGCCAGAGATGATCTCTCTCG 583

151 AlaIleTrpIleAspCysGlyIleIleIleAlaArgIleUtrPheSerProAl 167
 584 TTTCGCTGTGCTTCATAGGCCATATTCGAATGTGAGAAAGACCGTT 633
 167 aPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArgIleGlyAsnArg 184
 634 CTTCCTATGCGAACAAATCATGTGATCGGAACAGACACATGAAATGACCTT 683
 184 eRheTrpAlaAsnAsnHisCysIleGlyThrAspLeuAsnAsnArgAsnPhe 200
 684 GTCTCCAAACACATGTGTGAGGAAGTGCATGCCATGTTCTCATGCTCGGA 733
 201 AlaSerIleHisTrpCysGluGluGluAlaSerSerSerSerSerGlu 217
 734 AACCTAGTGGACTTATATCTGTCAGAACCAAGAAAGTGAAGGCACTGG 783
 217 UThrTrpCysGlyLeuTrpProGluSerGluProGluValIleValAla 234
 784 CTAGTTCTTGAGAGAAGAAATATCAACCCAGATTAAAGCATACATCAGCATG 833
 234 IAspPheLeuArgArgAsnIleAsnGlnIleValIleValIleValIleVal 250
 834 CATTCATACCTCCAGCATATATGTTTCCATATTCATATACAGCAAGTAA 883
 251 HisSerTrpSerGlnHisIleValIlePheProTrpSerTrpThrArgSer 267
 884 AAGCAAGACCATGAGGAGACTGTCTGTATGAGCAGAGGAGGAGTTCGTG 933
 267 sSerIleAspHisGluGluLeuSerLeuValAlaSerGluAlaValAlaArg 284
 934 CTATTCAGAAACTAGTAAATATACAGATATACATGAGTGGCATGCTCA 983
 284 IAlleGluIleValTrpSerIleAsnThrArgTrpThrHisGluHisGlySer 300
 984 GAACCTTATACCTAGCTCTCTGAGAGTGGGAGAGATGATGATGATTT 1033
 301 GluThrLeuTrpLeuAlaProGlyGlyGlyAspAspTrpIleTrpAsp 317
 1034 GGGCATCAAAATATGCTTTCATCAACCAACCTGTAGAGAACCTTTGCG 1083
 317 uGlyIleIleValTrpSerPheThrSerAsnProProValGluValLeu 334
 1084 CGCTGTCTCTAAAA 1097
 334 roleUserSerLeuIleVal 338
 seq_name: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.AAB11458
 seq_documentation_block:
 ID AAB11459 standard: Protein; 246 AA.
 XX
 XX AAB11459;
 XX
 XX
 XX 01-MAR-2001 (first entry)
 XX
 DE Human brain carboxypeptidase B protein SEQ ID NO. 4.
 XX
 KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.
 XX
 OS Homo sapiens.
 XX
 PN WO200066717-A1.
 XX
 XX 09-NOV-2000.
 PD
 XX 01-MAY-2000; 2000WO-JP02878.
 PF
 XX 30-APR-1999; 99JP-0125169.
 PR

XX (MATS/) MATSUMOTO A.
 PA
 XX
 XX Matsumoto A;
 PI
 XX
 DR WPI: 2000-687534/67.
 XX
 PT Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 PS
 PS Disclosure: Page 74-75; 84pp; Japanese.
 CC
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 CC
 SO Sequence 246 AA:

alignment_scores: Quality: 1325.00 Length: 246
 Ratio: 5.408 Gaps: 0
 Percent Similarity: 99.593 Percent Identity: 98.780

alignment_block:

US-09-980-881-1 x AAB11459 ..

Align seq 1/1 to: AAB11459 from: 1 to: 246

360 GCGTCCGCGATGCTATGAGACATACACATCAATGAAATCTATTC 409
 1 AlaseraSerlyrtyrGlnGlnItyrHisSerLeuasnGlnItyrSe 17
 410 TTGGATGAAATTTTAACTGAGAGCATCCGATATGCTTCAAAAATCC 459
 17 rTPrIleGluPheIleThrGluAlaGHisProAspMetLeuThrIysIleH 34
 460 ACATTTGGATCCCTCATTTGAGAAATGACCACTGTATGTTTAAAGTTTCT 509
 34 IsIleGlySerSerPheGlnIlystyrProleuTyValIleuIysValSer 50
 510 GGAAGAAACAAACAGCCAAAATGCCATATGATGATGACTGTGAAATGCA 559
 51 GILySGlnGlnThrAlaIlyAsnAlaIleTPrIleAspCysGlyIleH 67
 560 TGGCAGAGAAATGATCTCTCTGCTTCTGCTTGGTTCATAGGCCATA 609
 67 sAlaArgGluTPrIleSerProAlaPheCysLeuTPrPheIleGlyHisA 84
 610 ATCGAATGTGGAGAAAGACCGTTCTTTCTATGGCAACATCATTTGCATC 659
 84 snArgMetTPrArgIlyAsnArgSerPheTyAlaAsnAsnHisCysAlle 100
 660 GGAACAAACCTGGAATAGCACTTTGTCTCCAAACACTGGGTGAGGAAG 709
 101 GILyThrspLeuAsnArgAsnPheAlaSerTyHisIlyTPrCysGlnGluG 117
 710 TGCATTCAGTTCCTCATGCTCGAAGACCTACTGTGACTTATATCTGAGT 759
 117 yAlaSerSerSerSerCysSerGluTPrIlyrCysGlyLeuTyTProGluS 134
 760 CAGACACAGAAATGAGGACAGCTGGCTAGTTTCTTGAGAAAGAAATATCAAC 809
 134 eArgIUProGluValIlyAlaValAlaSerPheLeuArgArgAsnIleAsn 150
 810 CAGATTAAAGCATATCATGATGATTCATTAATCTCCAGCATATAGTGT 859

151 GlnIleuYsAlaTyrlIleSerMetHisSerTySerGlnHisIleValPh 167
 860 TCCATATTCCTATACAGAGTAAGAAAGACCATGAGCAATCTGTCTC 909
 167 eProTySerTyThrIlyrArgSerIlySerIlyAspHisIleGluLeuSerL 184
 910 TAGTAGCCAGTAGAAGACGTTGCTGCTATTGACAAATACTAGTAAATAC 959
 184 euValAlaSerGluValAlaValAlaGAlaIleGluIlyThrSerIlyAsnHr 200
 960 AGGTATACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCGAGAG 1009
 201 ArgTyThrHisGlyHisGlySerGluThrPheuTyrLeuAlaProGlyG 217
 1010 TGGGACGATGGATCTATGATTTGGGATCAAAATATCGTTTACATCA 1059
 217 yGlyAspAspTrpIleTyThrAspLeuGlyIleIlyTySerPheThrSerA 234
 1060 ACCCACCTGTAGAGAAAGCTTTGCCGCTGTCTCTAATA 1097
 224 snProProValGluIlyIlyLeuProLeuSerLeuIlyS 246

seq_name: /STD1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: AAB58129

seq_documentation_block:

ID AAB58129 standard; Protein; 211 AA.

AC AAB58129;

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polypeptide sequence SEQ ID 467.

XX Human: lung cancer associated protein; neuroprotective; cyrostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnertary;
 KW gastrointestinal; nephrotoxic; antineoplastic; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease.

XX Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05918.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI: 2000-587514/55.

DR N-PSDB; AAF18005.

PT Lung cancer associated gene sequences, referred to as lung cancer

PT such as lung cancer -

PS Claim 11: Page 944-945; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytosolic; cardioactive;
 CC immunomodulatory; muscular active general; vulnertary; gastrointestinal
 CC general; nephrotoxic; antineoplastic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer.

Percent Similarity: 64.678 Percent Identity: 33.890

alignment_block:

US-09-980-881-1 x AAM06172 ..

Align seg 1/1 to: AAM06172 from: 1 to: 415

```

33 CTTCGACGCTCTGACCATGTTCTCTCTGACGACGACATGCTTCGCG 82
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGly 18
83 G... TTCCAGAGTGGCCAGTTCTAGCTGCTCTCTAGAACCTCTAGCC 129
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 UHSphGluGlyGlyValPheArgValAsnValGluAspGluSer 35
130 AAGTTCAAGTCTACAGAAATCTTACTACACATATGATGTTCTCTG 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 IsIleAsnIleIleArgIleuValaSerThrThrGlnIleAspPhe 51
180 CAGCCGTTACAGCTGACCTTATTGTGAGAAAAACAGATCTCATTT 229
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 LysProAspSerValThrGlnIleLysProHisSerThrValaAsp 68
230 TGTAAATGACATCGATGTCGACATGGAAGCCCATTTAAATGACCG 279
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 GValLysAlaGluAspThrValThrValGluAsnValLeuLysGln 85
280 GAATTCATGACGATGTTCTGCTGCGACGACGTGAAGATCTTATCA 329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 IuLeuGlnIlyrLysValLeuIleSerAsnLeuArgAsnValValG 101
330 CAGATTTCCAAAGACAGACAGTCCGCGACGCTCCGATCTACTATGA 379
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 GlnPheAspSerArgVal..... ArgAlaThrGlnHisSerTyrG 115
380 ACAGTATCAGCTACTAAATGAAATCTATCTTGATAGATTTAACTG 429
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 ULysTyrAsnLysTrpGlnThrIleGlnAlaIleThrGlnIleVal 132
430 AGAGCATTCCTGATATGCTTACAAAATCCATTTGATCTCTATTTAG 479
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 HrgIuAsnProAlaLeuIleSerArgSerValIleGlyThrThrPhe 148
480 AAGTACCACCTATATGTTTAAAGTTTCTGAAAACAAACAGCCAA 529
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 GlyArgAlaIleTyrLeuLeuLysVal... GlyLysAlaGlyGln 164
530 AAATGCCATATGATTTGACTGTGAATCCATGCCAGAGATGATCTTC 579
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 sProAlaIlePheMetAspCysGlyPheHisAlaArgGluTrpIle 181
580 CTGCTTCTGCTGTTGTTGTTA..... 602
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 RoAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGly 197
602 ..... 602
198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeu 214
603 .....GGCCAT.....AATGCAATGT 618
214 ovalleuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgP 231
619 GGAGAAAGAACGTTCTTCTATGAGAACATTCATTCATGAGAACAC 668
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 rPalGlySerThrArgSerThrHisThrGlySerCysIleGlyThr 247
669 CTGAATAGCAACTTGTCTCAAAACACTGTGTGAGGAAGGTGATCC 718
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 ProAsnArgAsnPhe... AspAlaGlyTrpCysGluIleGlyAla 263
719 TTCCCTCATGCTCGGAACCTACTGTGAGCTTTATTCCTGATCAGCA 768
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

263 gAsnProCysAspGluThrTyrCysGlyProAlaAlaGluSerGluLys 280
769 AAGTGAAAGCAGTGGCTAGTTCCTTGAGAGAAATATCAACAGATTAA 818
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
280 IuThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296
819 GCATACATCAGCATGATTCATCTACTCCGATATAGTGTTCATATTC 868
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
297 AlaTyrLeuThrIleHisSerTyrSerGlnMetIleTyrProTyrSe 313
869 CTATACAGCAGTAAGCAAGAACCATGAGGACTCTCTAGAGCCA 918
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 rTyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAla 330
919 GTGAGCAGATTCGTCTATTCACAAACTAGTAATAAATACAGGTATACA 968
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 yAlaIleThrValLysGluLeu... AlaSerLeuHisGlyThrLysTyr 345
969 CATGGCCATGGCTCAGAAACCTTATACCTGCTCTGAGGTGGGAGCA 1018
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 TyrGlyProGlyAlaThrThrIleTyrProAlaAlaGlySerAsp 362
1019 TTGGATCTATGATTTGGGCATCAATATTCGTTTACATC..... 1057
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 rTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPheGluLeuArg 379
1058 .....AAACCCACC 1066
379 sPThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395
1067 TGTAGAGAGCTTTTGCCGCTGTCTTAATAATAGCTTGACATGATTAG 1116
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
396 CysGlnGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeu 412
1117 GAATGT 1123
412 ..... 414
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:AAW74476
seq_documentation_block:
ID AAW74476 standard; protein; 415 AA.
XX
AC AAW74476;
XX
DT 20-NOV-1998 (first entry)
XX
DE Human pancreatic carboxypeptidase protein sequence.
XX
KW Human; pancreatic carboxypeptidase B; insulin; protein sequencing;
  KM prodrg therapy.
XX
OS Homo sapiens.
XX
WO9835988-A1.
XX
PD 20-AUG-1998.
XX
PE 10-FEB-1998; 98MO-GB00415.
XX
PR 29-OCT-1997; 97GB-0022727.
PR 14-FEB-1997; 97GB-0003104.
PR 18-OCT-1997; 97GB-0022003.
XX
PA (ZENE ) ZENEPHA (ZENE ) ZENEPHA LTD.
PI Edge MD;
XX
DR WPI; 1998-467168/40.
DR N-PSDB; AAVAL1795.
XX
PT New modified pro-domain of carboxy-peptidase B - enhances expression
  of co-expressed proteins for production of recombinant

```

PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
 enzyme producing therapy

PS Example 1; Page 54-55; 83pp; English.

CC The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
 CC modified pro-domain of CPB on a separate gene to enhance recombinant
 CC expression. This process can be used to produce recombinant CPB in
 CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
 CC insulin production and protein sequencing, while its fusions with
 CC antibodies are useful in antibody-directed enzyme producing therapy. The
 CC modified pro-domain provide increased yields of recombinant CPB, possibly
 CC by protecting the C-terminus against enzymatic degradation or by
 CC increasing intracellular trafficking.

XX Sequence 415 AA;

alignment_scores:
 Quality: 651.50 Length: 419
 Ratio: 2.404 Gaps: 8
 Percent Similarity: 64.678 Percent Identity: 33.890

alignment_block:
 US-09-980-881-1 x AAW74476 ..

Align seg 1/1 to: AAW74476 from: 1 to: 415

```

33 CTTCGATGCTTGTACCATTTCTCTCTGTGACGACGACGCTTCGCG 82
   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisIsglyGly 18
83 G...TTCCAGAGTGGCCAGTCTAGCTGCTTCTAGAACCTTAGGC 129
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 uHisPheGluGlyGlyValPheArgValAsnValGluAspGluAsn 35
130 AAGTCAAGTCTACGAATCTTACCAACAATAGATGATGCTTCCTG 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 IsIleAsnIleIleArgGluLeuAlaSerThrThrGlnIleAspPhe 51
180 CAGCCGCTACAGCTGACCTTATGTGAAGAAACAAGTCCATTTT 229
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 LysProAspSerValThrGlnIleLysProHisSerThrValAsp 68
230 TGTAAATGATCTGATGTCGACAATGTGAAGCCCATTTAAATGAG 279
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 GValLysAlaGluAspThrValThrValGluAsnValLeuLysGlu 85
280 GAATTCATGACAGTGTCTGTGCTGACGACGAGAGATCTTATTC 329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 IuLeuGlnIlyrLysValIleuIleSerAsnLeuArgAsnValVal 101
330 CAGATTTCCAGACGACAGCTGACGCGCCGACCTCCGATGCTACT 379
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 GlnPheAspSerArgVal.....ArgAlaThrGlyHisSerArg 115
380 ACAGTTCACATCTCAATGAATCATCTTGTGATAGAAATTTATAC 429
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 uLysTyrAsnLysTyrGluThrIleGluAlaTyrThrGlnGlnVal 132
430 AGAGGACCTGTATGCTTACAAAATCCACATGTGATCCCTTTGAG 479
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 hGluAsnProAlaIleuIleSerArgSerValIleGlyThrThrPhe 148
480 AAGTACCCACTGTATCTTAAAGTTCTGAAAAAACAACAAACACCA 529
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 GlyArgAlaIleTyrLeuLeuLysVal...GlyLysAlaGlyGln 164
530 AATGCCATATGATGATGCTGGAATCCATGCGACAGATGATCTTC 579
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 sProAlaIlePheMetAspCysGlyPheHisAlaArgIuThrIle 181
580 CTGCTTTCTGCTTGTGTCATTA..... 602

```

```

181 roAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197
602 ..... 602
198 GluIleGlnValThrGluLeuAspLysLeuAspPheTyrValLeuPr 214
603 .....GCCAT.....AATCGAATGT 618
214 ovalLeuAsnIleAspLysTyrIleTyrThrThrThrLysSerArgPhe 231
619 GAGAAAGAACCGTCTTCTTATGCGACATCATCTTCATCGGACAGAC 668
231 rPArgLysTrpArgSerThrHisThrGlySerCysIleGlyThrAsp 247
669 CTGAATAGCAACTTGTCTCCAAACACTGGTGTGAGAGAGTGCATCAG 718
248 ProAsnArgAsnPhe...AspAlaGlyTrpCysGlnIleGlyAlaSer 263
719 TTCCCTATGCTCGAAGAACCTACTGTGACTTATTCCTGACTGACAG 768
263 gasnProCysAspGluThrTyrCysGlyProAlaAlaGluSerGluLys 280
769 AAGTGAAGCAGTGGCTAGTCTTGTGAGAAGAATATCAACGATTTAA 818
280 IuThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerIleLys 296
819 GCATTCATCAGATGATCATCTACCTCCGACATATAGTGTTCATATTC 868
297 AlATyrIleuThrIleHisSerTyrSerGlnMetMetIleTyrProLys 313
313 rTyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAla 330
919 GTAGACAGTTCGCTATGTACAAAAGTGTAAATATACAGATATACA 968
330 ysaIleThrValLysGluLeu...AlaSerLeuHisGlyThrLysTyr 345
969 CAGGCCATGCTCAGAAACCTTATACCTGCTGCTGAGAGTGGAGCA 1018
346 TyrGlyProGlyAlaThrThrIleTyrProAlaAlaGlyLysSerAs 362
1019 TTGATCATATGATTTGGGACATCAATATTCGTTACATC..... 1057
362 pTrpAlaTyrAspGlnGlyLeuArgTyrSerPheThrPheGluLeuArg 379
1058 .....AAACCACC 1066
379 sPThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395
1067 TGTAGAGAAGCTTTTCCGCTGTCTTAAAAATAGTTGGACGTGCATTA 1116
396 CysGluGlnThrPheLeuAlaIleLysTyrValAlaSerTyrValLeu 412
1117 GAATGTT 1123
412 uHisLeu 414

```

seq_name: /SIDS1/gcgdelta/geneseq/geneseqp_emb1/AA1996.DAT:AAW06175

seq_documentation_block:

ID AAW06175 standard; Protein; 424 AA.

XX AAW06175;

XX 17-FEB-1997 (first entry)

DE PROCPB with Peib leader sequence.

XX ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPPT;

KW mustard-ribonuclease; antibody directed enzyme producing therapy;

KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;

reduced immunogenicity; primer: PCR; polymerase chain reaction; HP-RNase; Fd; Ftab'2; PeIB; leader; human carboxypeptidase B.

Synthetic.

Key	Location/Qualifiers
Peptide	1..22
Peptide	/label= PeIB_leader_sequence
Peptide	23..117
Protein	/label= pro_sequence
	118..424
	/label= mature_HCPB

WO9620011-A1.

04-JUL-1996.

21-DEC-1995; 95WO-GB02991.

16-AUG-1995; 95GB-0016810.

23-DEC-1994; 94GB-0026192.

(ZENEC) ZENECAL LTD.

Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW, Henham JF, Hennequin LRA, Marsham PR, Rabin BR, Slater AM, Tarragona-Fiol A, Taylorson CJ;

MP1; 1996-321650/32.

N-PSDB; AAT42506.

Two component system for anti-tumour therapy - comprising targeting moiety linked to mutated enzyme which can transform an anti-neoplastic produg

Reference Example 18; Page 140-142; 182pp; English.

A two-component system for anti-tumour therapy comprises a targeting moiety linked to a mutated enzyme which can transform an anti-neoplastic produg. The system is based on antibody directed enzyme produg therapy (ADEPT) using non-naturally occurring mutant forms of host enzymes, pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).

Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can be used. The present sequence is that of proHCPB with a PeIB leader sequence encoded by PIC11738.

Sequence 424 AA;

alignment_scores:

Quality:	644.00	Length:	401
Ratio:	2.458	Gaps:	7
Percent Similarity:	65.337	Percent Identity:	33.915

alignment_block:

US-09-980-881-1 x AAM06175 ..

Align seg 1/1 to: AAM06175 from: 1 to: 424

```

84 TTTCAGAGTGGCCAGTTCTAGCTCTCTCTAGAACCTCTAGGCAAGT 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 PheGIGlyGlyValPheArgValAsnValGlnAspGlnAsnHisI 45
134 TCAAGTCTACAGAACTTACTACAACTATGAGATTGTTCTGCGAGC 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 easnIleIleArgGlnUeuAlaSerThrThrGlnIleAspPheTrpLysP 62
184 CGGTAAAGCTGACCTTATTTGAGAGAAAAACAAGTCCATTTTGTGA 233
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 roAspSerValThrGlnIleLysProHisSerThrValAlaPhePheVal 78
234 AATGCAATCTGATGTCGACATGTGAAGCCCATTTAAATGTGAGCGGAA 283
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

79 LysAlaGlnAspThrValThrValGlnAsnValLeuLysGlnAsnGlnLeu 95
284 TCCATGAGTGTCTTGTGCGAGACGTGAGATCTTATTCACAGCAGA 333
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 uGlnTyLysValLeuIleSerAsnLeuArgAsnValValGlnLacInP 112
334 TTTCGAACGACACAGTCAGCCCGCGACCTCCGATGCTATGAAACAG 383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
112 heAspSerArgVal.....ArgAlaThrGlnHisSerTyrGlnLys 125
384 TATCACTCACTAAATGAATCTATCTTGTGATAGATTTTATCTGAGAG 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
126 TyrAsnLysTrpGlnThrIleGlnAlaThrPhrGlnGlnValAlaThrG 142
434 GCATCTGTATATGCTTACAAAATCCACATTGATCTCATTTGAGAGT 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 uAsnProAlaLeuIleSerArgSerValIleGlyThrThrPheGlnGly 159
484 ACCCATCTATGTTTAAAGTTTCTGGAAGAAACAACAGCCAAAAT 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159 rGlnAlaIleTyrLeuLeuLysVal...GlyLysAlaGlyGlnAsnLysPro 174
534 GCCATATGATGATGCTGTGGAATCCAGCCAGACATGATCTCTCTGC 583
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 AlaIlePheMetAspCysGlyPheHisAlaArgValTrpIleSerProAl 191
584 TTCTGCTGTGGTTCATA..... 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 aPheCysGlnTrpPheValArgGlnAlaValArgThrTyrGlyArgGln 208
602 ..... 602
208 IeGlnValThrGlnLeuLeuAspLysLeuAspPheTyrValLeuProVal 224
603 .....GGCCAT.....AATGAAATGTGAG 622
225 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrp 241
623 AAGAACCGTCTTCTTATGCGAACAAATCATGTCATGGAACAGCTGA 672
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 glyThrArgSerThrHisThrGlySerSerCysIleGlyThrAspProA 258
673 ATGACAACTTTGTCCTCCAAACACTGCTGTGAGAGGTCATCCAGTTC 722
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
258 snArgAsnProe...AspAlaGlyTrpCysGlnIleGlyAlaSerArgAsn 273
723 TCATGCTCGAAACCTACTGCTGACTTTATCTAGTCAGACAGCAAGT 772
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 ProCysAspGlnThrTyrCysGlyProAlaAlaIleSerGlnLysGln 290
773 GAAAGCAGTGGCTAGTTCTTGAGAGAAATATACACAGATTAAAGCAT 822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 rLyAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLysAla 307
823 ACATCAGATGATTCATCTACCTCCAGCATATAGTGTTCATTTCTCAT 872
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 yLeuThrIleHisSerTyrSerGlnMetIleTyrProTyrSerTyr 323
873 ACAGCAAGTAAAGAACCAAGCATGAGAACTGCTAGTAGACGAGGA 922
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 AlaTyrLysLeuGlyGlnAsnAsnAlaGlnLeuAsnAlaLeuAlaLysAl 340
923 AGCAGTTCGTGCTATTTGACAAAACTAGTAAATAACAGGTATACAGATG 972
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
340 aThrValLysGlnLeu...AlaSerLeuHisGlyThrLysTyrThrGlyG 356
973 GCCATGGCTCAGAAACCTTATACCTAGCTCTGAGAGTGGGAGAGATTGG 1022
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 LyrProGlyAlaThrThrIleTyrProAlaAlaGlyLysSerAspAspTrp 372
1023 ATCATGATTTGGGCAATCAATATCTTACATC..... 1057
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 AlaTyrAspGlnGlyIleArgTyrSerPheThrPheGlnLeuArgAspThr 389

```

1058AAACCCACCTGTA 1070
389 rGLyArTYrGLyPheLeuLeuProGLuSerGLnIleArGLaIaThrCYag 406
1071 GAGAGCTTTGCCGCTGCTCTAAATGCTTGCAATGCTCAATAGAGT 1120
406 IuGLuThrPheLeuAlaIleLysTYrValAlaSerTYrValLeuGLnHis 422
1121 GWT 1123
423 Leu 423
seq_name: /SIS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.AAR75131
seq_documentation_block:
ID AAR75131 standard; Protein: 404 AA.
XX AAR75131;
AC
XX 13-MAR-1996 (first entry)
DT
XX Porcine Tyr-His-Met Procarboxypeptidase B.
DE
XX Procarboxypeptidase B: carboxypeptidase B; Pichia; pCPB;
KM human serum albumin; premating factor alpha; mating factor alpha;
KW pCPBB.
XX
XX Sus scrofa.
OS
XX WO9514096-A1.
PD 26-MAY-1995.
XX
XX 16-NOV-1994; 94WO-US13142.
PF
XX 16-NOV-1993; 93US-0153258.
PR
XX (ELIL) LILLY & CO ELI.
PA
XX Fayerman JT, Greenen DP, Hersherberger CL, Larson JL;
PI Steiner JL, Zhang H;
PI
XX WPI: 1995-200386/26.
DR N-PSDB; AAQ90600.
XX
XX DNA encoding porcine carboxypeptidase B - used for transforming
PT host cells, partic. Pichia species, for prodn. of the enzyme
PI
XX
PS Example 1: Page 16-17; 34pp; English.
XX
XX The porcine carboxypeptidase B coding sequence can be place in a
CC bacterial or pref. Pichia yeast expression vector. The expression
CC vector further comprises the signal peptide of either human serum
CC albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC (designated pF0489 - NRRL B-21028); mating factor alpha (designated
CC pF0474 - NRRL B-21032) or the porcine pCPBB signal peptide;
CC (designated pLGD27 - NRRL B-21027). The method can be used for
CC producing large amounts of porcine carboxypeptidase B and when
CC produced in Pichia yeast, the protein does not need solubilisation
CC or folding. The produced enzyme is then used for pref. cleaving
CC basic residues from the carboxy terminus of proteins.
XX
XX
SQ Sequence 404 AA;
alignment_scores:
Quality: 640.50 Length: 361
Ratio: 2.680 Gaps: 6
Percent Similarity: 66.205 Percent Identity: 37.673
alignment_block:
US-09-980-881-1 x AAR75131 ..

Align seg 1/1 to: AAR75131 from: 1 to: 404
84 TTCCAGATGGCCCAATGCTAGCTGCTTCCTGCAAGCTGAGCAAGT 133
10 PheGLuGLyGLuLysValPheArgValAsnValGLuAspGLuAsnPII 26
134 TCAAGTTCTACAGAACTTACTACAAATATGATGATGCTGCTGAGC 183
26 eSerLeuLeuHisGLuLeuAlaSerThrArgGLnIleAspPheTrpLys 43
184 CGGTACACAGCTGATTTATGTCAGAAAAAACAGTCCATTTTGTGA 233
43 rAspSerValThrGLnIleLysProHisSerThrValAspPheArgVal 59
234 AATGATCTGATGTCAGCAATGCAAGGCCAATTTAAATGTCAGCGAGT 283
60 LysAlaGLuAspIleLeuAlaValGLuAspPheLeuGLnHisGLuLe 76
284 TCCATGACAGTGTCTGCTGCGACAGCTGGAGATCTTATTCACAGCAG 333
76 uGLnTYrGLuValLeuIleAsnHisLeuArgSerValLeuGLuAlaGLn 93
334 TTTCACACGACAGTCAGCCCCGAGCTCCGATCTGACTATGACAG 383
93 hAspSerArgVal.....ArgThrThrGLyHisSerTYrGLuLys 106
384 TATCACTACATGAATGAAATCTATTCTGATAGATATTTATACAGAG 433
107 TYrAsnAsnTrpGLuThrIleGLuAlaTrpTrpLysGLuValIleSer 123
434 GCATCTGATATGCTTACAAAAATCCATGGATCCATTTGAGAGAGT 483
123 uAsnProAspLeuIleSerArgThrAlaIleGLyThrPheLeuGLyA 140
484 ACCCACTATGTTTAAAGTTCTCGAAAGAACAAACAGCCAAAT 533
140 sAsnIleTYrLeuLeuLysVal...GLyLysProGLyProAsnLysPro 155
534 GCCATATGATGATGACTGTGGAATCCATGCCAGAGATGAGCTCCCTC 583
156 AlaIlePheMetAspCysGLyPheHisAlaArgIleTrpIleSerHisAl 172
584 TTTCTGCTTGCTGATCA..... 602
172 aPheCysGLnTrpPheValArgGLuAlaValLeuThrTYrGLyTYrGLu 189
602 602
189 eRHisMetThrGLuPheLeuAsnLysLeuAspPheTYrValLeuProVal 205
603GCCAT.....AATCGAATGGAG 622
206 LeuAsnIleAspGLyTYrIleTYrThrTrpTrpLysAsnArgMetTrpArg 222
623 AAGAAGCGTTCTTCTATCGAACATATGATCGAGACAGACCTGA 672
222 gLysTrpArgSerThrAsnAlaGLyThrThrCysIleGLyThrAspPro 239
673 ATAGCAACTTGTCTCCAAACACTGCTGAGAGAGTGCATCCAGTTCC 722
239 sNArgAsnPhe...AspAlaGLyTrpCysThrThrGLyAlaSerThrAsp 254
723 TCATGCTCGAAACCTACTGAGCTTATCCGATCGACAGACCAAGT 772
255 ProCysAspGLuThrTYrCysGLySerAlaAlaGLuSerGLuLysGLuThr 271
773 GAGGACAGTGCCTAGTTCTTGAGAGAAATATCAACGATTAAGCAT 822
271 rLysAlaLeuAlaAspPheIleArgAsnAsnLeuSerIleLysAla 288
823 ACATGACATGATCATCTATCCAGCATATAGTGTTCATATTCCTAT 872

```

288 yIleuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 304
873 ACACGAAGTAAAGCAAGACCATGAGCAACTGCTCTAGTACGCAAGTGA 922
      ::::::::::::::::::::
305 AspTyrIleuProGluAsnAsnIleuLeuAsnIleuAlaTyrSAI 321
923 AGCAGTTCGCTATGTGACAAACTAGTAAATAACGAGTATACATG 972
      |||||:::
321 aAlaValIleuGluLeu...AlaThrLeuTyrGlyThrLysTyrThrTyrG 337
973 GCCATGCGCTGACAAACCTTATACCTGCTCGGAGGTGGGACGATGG 1022
      |||||:::
337 LyrProGlyAlaThrThrIleTyrProAlaIleGlyGlySerAspAspTyr 353
1023 ATCTAGATTTGGCGCATCAATATTCGTTACA 1055
      |||||
354 AlaTyrAspGlnGlyIleLysTyrSerPheThr 364

seq_name: /STDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA66562

seq_documentation_block:
ID      AA66562 standard; Protein; 324 AA.
XX
AC      AA66562;
XX
DT      22-OCT-2001 (first entry)
XX
DE      Bothops jararaca carboxypeptidase homologue protein.
XX
KW      secreted metalloprotein; carboxypeptidase; immunomodulatory;
KW      antiparasitic; nootropic; anticonvulsant; neuroprotective;
KW      osteoporosis; antidiabetic; cerebroprotective; vasotropic; vulnary;
KW      antibacterial; antiviral; antifungal; antiinflammatory; antidiabetic;
KW      antineoplastic; cytostatic; antiallergic; haemostatic; infection;
XX      cancer.
XX
OS      Bothops jararaca.
XX
PN      WO200157265-A1.
XX
PD      09-AUG-2001.
XX
PF      05-FEB-2001; 2001WO-US03783.
XX
PR      03-FEB-2000; 2000US-0496914.
PR      27-APR-2000; 2000US-0560875.
PR      29-SEP-2000; 2000US-0676135.
XX
PA      (HYSE-) HYSEQ INC.
XX
PI      Boyle BJ, Mize NK, Arterburn MC, Yeung G, Tang YT, Liu C;
PI      Drmanac RT, Wang M, Chen L, Yang Y;
XX
DR      WPI; 2001-465705/50.
XX
PT      Novel metalloproteinase peptidase-like polypeptide is useful in treating
PT      neurodegenerative, immune, etc. disorders
XX
PS      Example 3; Fig 1; 146pp; English.
XX
CC      The invention relates to an isolated polypeptide comprising an amino
CC      acid sequence of 374, 41, 14, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
CC      354, 165, 374, 354, or 41 amino acids, fully defined in the
CC      specification, or its translated protein coding portion. The polypeptide
CC      coding portion, extracellular portion or active domain. The polypeptide
CC      is a human secreted metalloproteinase-like polypeptide. It
CC      is useful for treating mammalian diseases including:
CC      neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
CC      osteoporosis, osteoarthritis, disorders linked to reduced tissue
CC      growth, Huntington's disease, amyotrophic lateral sclerosis, stroke,
CC      reperfusion, wound healing, infectious diseases, immune and autoimmune
CC      disorders such as severe combined immunodeficiency, rheumatoid
CC      arthritis, Guillain-Barre syndrome, insulin dependent diabetes

```

```

CC      mellitus, graft-versus-host disease and allergy, cancer, and
CC      haemophilia. The polypeptide is also useful in medical imaging, as an
CC      immunogen to generate antibodies, as a nutritional source or
CC      supplement, and as a contraceptive. The present sequence is
CC      a carboxypeptidase homologue protein of Bothops jararaca.
XX
SQ      Sequence      324 AA;

alignment_scores:
      Quality: 636.00      Length: 329
      Ratio: 2.917      Gaps: 6
Percent Similarity: 66.261      Percent Identity: 40.122

alignment_block:
US-09-980-881-1 x AA66562 ..

Align seg 1/1 to: AA66562 from: 1 to: 324

267 TTAAATGTGACGCGAATTCATGACAGTGTCTTGTGCGACAGCGTGAAGA 316
      |||||
2      LeuGlnGlnSerGlyLeuAsnTyrGluIleuLeuIleAspAsnLeuGlnAl 18
317 TCTTATTCACAGCAGATTCACAGACAGACAGTACGCCCGAGCCTCG 366
      ::::::::::::::::::::
18 aValLeuAspArgGlnLeuAspAsnIleAla.....ArgThrAlaG 32
367 CATGCTACTATGAACAGTATCAGTATCAGTAAATGAATCTATCTGATA 416
      ::|||
32 LyrTyrAsnTyrGluLysTyrAsnSerTrpGluLysIleAspAlaThrPhe 48
417 GAATTTATTAAGTACAGAGCATCCTGATATGCTTACAAATTCACATGG 466
      |||||
49 AlaAspIleAlaAsnGlnLeuAsnProSerLeuValSerArgLeuGlnIleG 65
467 ATCCGATTTGAGAAAGTACCCACTATGTTTAAAGTTTCTGGAAGA 516
      ::::::::::::::::::::
65 YThrThrPheGlnGluArgProMetProLeuLeuLysVal...GlyLysP 81
517 AACAAACAGCCAAATATGCAATGATGATGATGATGATGATGATGATGAT 566
      ::|||
81 roGlyValAsnLysLysAlaIlePheIleAspCysGlyPheIleAlaArg 97
567 GAATGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
      |||||
98 GluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValAr 114
602 .....
114 gThrTyrGlyLysGluThrIleMetThrGlnLeuLeuAsnLysLeuAsp 131
603 .....GCG 605
131 heTyrIleLeuProValLeuAsnIleAspGlyTyrValTyrSerTyrLys 147
606 CATAATGCAATGTGAGAAAGAACCGTCTTCTTATGGAACATCATG 655
      ::::::::::::::::::::
148 GlnSerArgMetTrpArgLysThrArgSerValAsnAlaGlySerThrCy 164
656 CATGGAACAGACCTGAATAGCACTTGTCTCCAAACACTGCTGTGAGG 705
      |||||
164 sIleGlyThrAspProAsnArgAsnPhe...AspAlaIleTrpCysSerV 180
706 AAGGCGATTCAGTTCATGCTGCGAAGCTGAGTGGATCTATATCT 755
      |||||
180 aGlyAlaSerArgAsnProCysSerGluThrTyrCysGlySerLysPro 196
756 GAGTCAGAACCAAGAGTGAAGGCACTGCTTCTTGAAGAAATAT 805
      |||||
197 GlnSerGluLysGluThrTyrAlaLeuAlaAspPheIleArgArgAsnAr 213
806 CAACAGATTAAGCATATCATCATGATCATCATCATCATCATCATCATCAT 855
      ::|||

```



```

603 .....GGCCATATCGAATGTGGAGA 623
238 heasnValaspGlyTyrHisPheSerTrpThrAsnAspArgPheTrpArg 254
624 AAGAACGGTTCTTCATGCAACATCATGATCGGACAGACCTGAA 673
255 LysThrArgSerArgAsnSerArgPheArgCysArgGlyValAspAlaAs 271
674 TAGCAACTTGTCTCAAAACAGTGTGAGGAAGTGATCCAGTTCT 723
271 nArgAsnTrpArgValLys...TrpCysAspGluGlyAlaSerMetHisP 287
724 CATGCTGGAAACCTACTGTGAGATTATCTGTAGTCAGAACCGAAGTG 773
287 roCysAspAspTrpTyrCysGlyProPheProGluSerGluProGluVal 303
774 AAGCAGGCGGTACTTCTTGAGAAAGAAATATCAACGATTAAGCTA 823
304 LysAlaValAlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTy 320
824 CATCAGCATGCATTCATCTCCAGCATATAGTGTTCATATTCCTATA 873
320 rLeuSerPheHisAlaTyrAlaGluMetLeuLeuTyrProTyrSerTyrL 337
874 CACGAGTAAAGCAAGACCATGAGACTGTCTTAGTACCAAGTCAA 923
337 ysrTyrAlaThrIleProAsnPheArgCysValGluSerAlaAlaTyrLys 353
924 GCAGTGTGCTATTGACAAACATAGTAAATACGAGGTATACACATNG 973
354 AlaAlaValAlaLeu...GlnSerValTyrGlyValAlaGlyTyrArg 369
974 CCATGCGCTCAGAAACCTATCTAGCTAGCTCGAGAGTGCGGAGCATGGA 1023
369 yProAlaSerThrThrLeuTyrValSerSerGlySerSerMetHisPrrpa 386
1024 TCTATGATTGGCATCAATATTCGTTAC..... 1054
386 lAtyTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThr 402
1055 .....ATCAACCCACCTGTAG 1071
403 GlyTyrPheGlyPheLeuLeuProGluMetLeuIleLysProThrCysTh 419
1072 AGAAGCTTTTGGCCGTCTCTCTAAATAGCTTGGCATGTCAATTAGCAATG 1121
419 rGluThrMetLeuAlaValLysAsnIleThrMetHisLeuLeuLysLys. 435
1122 TTTAATGCCCC 1132
436 .....CysPro 437
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW41414
seq_documentation_block:
ID AAW41414 standard; Protein; 716 AA.
XX
XX AAW41414;
AC
XX 02-JUN-1998 (first entry)
XX
XX PreproHCPB-linker-humanised 806.077 VH.
XX
XX Anti-CEA antibody; carcinoembryonic antigen; 806.077 Ab; cancer therapy;
XX cancer diagnosis; complementarity determining region.
XX
XX Synthetic.
XX
XX WO9742329-A1.
XX
XX 13-NOV-1997.
XX

```

```

PF 29-APR-1997; 97WO-GB01165.
XX
XX 14-FEB-1997; 97GB-0003103.
PR 04-MAY-1996; 96GB-0009405.
XX
XX (ZENEC ) ZENECA LTD.
XX
XX Copley CG, Edge MD, Emery SC;
XX
XX WPI; 1997-558987/51.
DR N-PDB; AAV17331.
XX
PT Anti-carcinoembryonic antigen antibody 806.077 Ab - used for
PT diagnosis and therapy of cancer
XX
PS Example 75; Page 184-187; 208pp; English.
XX
CC This sequence is preproHCPB-linker-humanised 807.077 VH, which is
CC an example of the antibody of the invention. The antibody is an anti-CEA
CC (carcinoembryonic antigen) antibody (preferably 806.077 Ab). Host cells
CC or transgenic organisms transformed with DNA encoding the antibody, are
CC used to make the antibody or conjugate. The conjugate is used in a
CC medicament suitable for intravenous administration. The conjugate can be
CC used for cancer therapy, selectively killing tumour cells. The antibody
CC can be used for in vivo or in vitro diagnosis of cancer.
XX
SQ Sequence 716 AA;

```

```

alignment_scores:
    Quality: 635.50      Length: 422
    Ratio: 2.345        Gaps: 8
Percent Similarity: 64.218      Percent Identity: 32.938

```

alignment_block:

US-09-980-881-1 x AAW41414 ..

Align seg 1/1 to: AAW41414 from: 1 to: 716

```

24 CTTTGCAGCCCTTGCAGCTCTGTATACCAATGTCTCTGTGACGACA 73
   :::::||||| ||||||| ::::::|||
1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisH1 17
74 TGTCTTGGCG...TTCAGAGTGGCCAAAGTTCTAGCTCTCTCTAGAA 120
   |:::::||||| ::::::|||
17 sGlyGlyGluHisPheGlyGlyGluValAlaPheArgValAsnValGluA 34
121 CCTCTAGCCAGTTCAAGTTCTACAGAACTTCTACACAACTATGAGATT 170
   :::::::||||| ::::::|||
34 sPgluAsnHisIleAsnIleIleArgGluLeuAlaSerThrThrGluIle 50
171 GTTCTCTGCGACCCGCGTAACAGCTGACCTTATTTGAGAAACAAACAGT 220
   :::::::||||| ::::::|||
51 AspPheTrpLysProAspSerValThrGluIleLysProHisSerThrVa 67
221 CCATTTTGTGTAATGCATCTGATGTGACAAAGTGAAAGCCATTGAA 270
   | ||| ||| :::::::||||| ::::::|||
67 lAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeuL 84
271 ATGTGAGCGGAATTCATGACAGTGTCTGTGCGGACGAGGTGGAAGATCT 320
   ::::: ||| :::::::||||| ::::::|||
84 ySglnAsnGluLeuGluIleTyrLysValLeuIleSerAsnLeuArgAsnVal 100
321 APTTCAACAGCAGATTTCCACGACGACAGCTCAAGCCCGACGCTCCGCAATC 370
   ::::::: ||| :::::::||||| ::::::|||
101 ValGluAlaGlnPheAspSerArgVal.....ArgAlaThrGlyH1 114
371 GTACTATGACAGATATCACTCACTAATGAATCTATTCTTGATAGAT 420
   ||| :::::::||||| ::::::|||
114 sSerTyrGluLysTyrAsnLysTrpGluThrIleGluAlaTrpThrGing 131
421 TTATTAAGTGAAGGCGATCTGATATGCTTACAAAAATCCACATGGATCC 470
   ::::::: ||| :::::::|||||

```

```
131 lnValAlaThrGluAsnProAlaLeuIleSerArgSerValIleGlyThr 147
471 TCATTGAGAGTACCCTATGTTTAAAGCTTCTGAAAGACACA 520
      ::::::::::::::::::::
148 ThrPheGluGlyArgAlaIleTyrIleuLeuLysVal...GlyLysAlaG1 163
521 AACAGCCAAAATGCGCATGTGATGATGCTGGATTCATGCCAGAGAA 570
      ||| ||||| ::::::::::::::::::::
163 yGlnAsnLysProAlaIleIlePheMetAspCysGlyPheHisAlaArgGlu 180
571 GCATCTCTCCCTGCTTCTGCTGTGCTCATTA..... 602
      ||||| ::::::::::::::::::::
180 rPileSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196
602 ..... 602
197 TyrGlyArgGluIleGlnValThrGluLeuAspLysLeuAspPheTyr 213
603 .....GCCCAT.....A 609
      ||:::
213 rValLeuProValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysS 230
610 ATCGAATGTGGAGAAAGACCTTCTTCTATGCCAACAATCATTCATC 659
      ::::::::::::::::::::
230 eArGpPheTrpArgLysThrArgSerThrHisThrLysSerSerCysIle 246
660 GGAACGAGCTGAATAGCACTTGTCTCCAAACACTGTGTGAGAGAG 709
      ||||| ||| ||||| ::|
247 GlyThrAspProAsnArgAsnPhe...AspAlaGlyTrpCysGluIleG1 262
710 TGCATCCAGTTCCTCATGCTGGAAACCTACGTGCACTTATCTGAGT 759
      ||||| ::| ||::| ||||| |||||
262 yAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGluS 279
760 CAGAACCAGAGTGAAGCGCAGTGGCTAGTTCCTTGAGAGAAGAATATCA 809
      ||||| ||::| ||||| ::||::| ||||| ::::::
279 eRGlLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295
810 CAGATTAAACATATACATGATCATCTTCACTCCAGCATATAGTGT 859
      ::::::::::::::::::::
296 SerIleLysAlaTyrIleuThrIleHisSerTyrSerGlnMetIleTyr 312
860 TCCATATTCCTATACAGAACTAAAGCAAGACATGAGGAAGTGTCTC 909
      ::::::::::::::::::::
312 rProTyrSerTyrAlaTyrLysLeuGlyLysAsnAsnAlaGluLeuAsnA 329
910 TAGAGCCAGTGAAGCAGTTCGTCTATGACAAAACCTAGTAAATAATAC 959
      ::::::::::::::::::::
329 laLeuAlaLysAlaThrValLysGluLeu...AlaSerLeuHisGlyThr 344
960 AGGTATACATGGCCATGGCTCAGAAACCTATATACCTAGCTCCTGGAG 1009
      ::::::::::::::::::::
345 LysTyrThrTyrGlyProGlyAlaThrThrIleTyrProSerAlaGlyThr 361
1010 TGGGAGCATGTTGATCTATGATTTGGGCAATATTCGTTTACATC.. 1057
      ::| ||||| ||||| ||||| ::|||
361 rSerLysAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPheC 378
1057 ..... 1057
378 luLeuAlaArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394
1058 AAACCCACCTGTAGAGAGCTTTGCCGCTGTCTCTAAATAGCTTGCA 1107
      ::| ||||| ::||| ||||| ::|||
395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTy 411
1108 TGTGATTAGGAATGTT 1123
411 rValLeuGluHisLeu 416
```

OM of: US-09-980-881-1 to: Issued_Patents_AA:* out-format : pfs

Date: Sep 18, 2002 4:39 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+g2p.model -DEV=x1h
-O=sgn2_1/USPROOPI/US09980881/runat_16092002_140039_8839/app-query.fasta_1.1655
-DB=Issued_Patents_AA -QFMT=fasta -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPTOL=0.000 -LOOEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAEXT=0.500
-DEL=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 -CGN1_1.14 -NCPU=6 -ICPU=3 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLIFY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-1
Query length: 1573
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 34.320000

```
score_list:
Sequence      Strd Orig      ZScore      EScore      Len      Documentation
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3 + 1885.00 3732.52 4.8e-201 423
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-3 + 1885.00 3732.52 4.8e-201 423
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-4 + 1885.00 3732.52 4.8e-201 423
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-08-869-057-2 + 1885.00 3732.52 4.8e-201 423
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-7 + 655.00 1276.60 3.0e-64 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-7 + 655.00 1276.60 3.0e-64 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-7 + 655.00 1276.60 3.0e-64 417
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-08-869-057-2 + 655.00 1276.60 3.0e-64 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-6 + 647.50 1261.62 2.1e-63 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-6 + 647.50 1261.62 2.1e-63 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-6 + 647.50 1261.62 2.1e-63 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-696-139-2 + 640.50 1248.01 1.2e-62 404
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-09-171-945-125 + 634.50 1229.42 7.5e-62 716
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-08-782-760-6 + 607.50 1185.29 5.0e-59 307
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:PCR-US96-00995-6 + 607.00 1181.35 6.5e-59 396
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-4 + 607.00 1181.35 6.5e-59 396
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-4 + 607.00 1181.35 6.5e-59 396
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-4 + 576.00 1122.43 1.6e-55 306
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-696-139-4 + 571.50 1105.41 7.1e-55 613
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-09-171-945-113 + 553.00 1072.92 6.7e-53 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-8 + 553.00 1072.92 6.7e-53 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-8 + 553.00 1072.92 6.7e-53 417
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-5 + 529.00 1024.95 3.2e-50 419
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-5 + 529.00 1024.95 3.2e-50 419
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-5 + 529.00 1024.95 3.2e-50 419
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-08-640-906-4 + 513.00 993.05 1.9e-48 417
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-09-395-936-4 + 513.00 993.05 1.9e-48 417
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-08-640-906-18 + 512.00 991.05 2.4e-48 417
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-09-395-936-18 + 512.00 991.05 2.4e-48 417
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-08-640-906-2 + 487.50 942.08 1.3e-45 419
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-09-395-936-2 + 487.50 942.08 1.3e-45 419
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-08-640-906-17 + 484.00 935.09 3.2e-45 419
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-09-395-936-17 + 484.00 935.09 3.2e-45 419
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-1 + 167.00 330.14 1.8e-10 37
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-277-540-1 + 167.00 330.14 1.8e-10 37
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-430-787A-1 + 167.00 330.14 1.8e-10 37
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-484-105-16 + 118.50 199.22 0.0002 706
/cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-08-484-105-16 + 118.50 199.22 0.0002 706
/cgn2_6/prodata/1/1aa/5B.COMB.pcp:US-08-782-760-3 + 100.50 186.45 0.0070 95
```

```
seq.name: /cgn2_6/prodata/1/1aa/5A.COMB.pcp:US-07-649-591B-3
seq_documentation_block:
: Sequence 3, Application US/07649591B
: Patent No. 5206161
:
: GENERAL INFORMATION:
: APPLICANT: Dennis Drayna and Daniel Eaton
: TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/649,591B
: FILING DATE: 19910201
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 689
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/266-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 423 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
:
: US-07-649-591B-3
:
: alignment_scores:
: Quality: 1885.00 Length: 423
: Ratio: 5.122 Gaps: 2
: Percent Similarity: 86.998 Percent Identity: 86.525
:
: alignment_block:
: US-09-980-881-1 x US-07-649-591B-3 ..
:
: Align seg 1/1 to: US-07-649-591B-3 from: 1 to: 423
:
: 18 ATGAAGCTTTCAGGCTTCAGCTTCCTTGAACCTTCTTGTGCA 67
: |||||||
: 1 MetLysLeuGlySerLeuAlaValLeuValProIleValLeuPheCysol 17
: 68 GGAGATGCTTCGGCTTCAGAGTGGCAAGTTCAGTCTTCCTCA 117
: |||||||
: 17 ucInIstValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
: 118 GAACCTTCAGGCAAGTTCAGATCTTACAGATCTTACATATGAG 167
: |||||||
: 34 rGThSerArgInValGlnValLeuGlnAsnLeuThrThrIryrGlu 50
: 168 ATTGTTCTGTGCGACCGGTAAACAGCTGACTTATTGAGAAAAACA 217
```

```

|||||
51 llevalleuiprGlnProvalThraAlaAspLeuIleValLysLysLysGI 67
218 AGTSCATTTTGTAAATGTCATGTGTCGACAATGGAAGCCATT 267
|||||
67 nvalnlspherhevalasnlaSeraspvalAspasnvalLysAlaHisL 84
268 TAAATGTGAGCGGAATTCATGCAAGTGTCTTGCGCAGAGCTGAGAT 317
|||||
84 euasnvalSerGIyleProcySerValleuIleuAlaAspValGIuAsp 100
318 CTATATCAACAGCATTTCCACAGACAGTCAGCCCGGAGCCCTCCG 367
|||||
101 leuileglnGlnInleSerasnAspThraValSerProAlaSerAl 117
368 ATGCTACTATGACAGTATCATCTCACTAAATGAATCTATCTTGATAG 417
|||||
117 aserGyTtYgLuGIInlyThraSerleuasnGluIleGySerTrpIleG 134
418 AATTATTAAGTAGAGGATCCGATATGCTTAACAATAATCCACATTTGA 467
|||||
134 lnphelelTrGIuAlaGnlsProAspMetleuThryLysIleHisIleGI 150
468 TCCTCATTTGAGAGTACCCGCTATGTTTAAAGTTTCTGGAAGA 517
|||||
151 SerSerPheGluLysTyrProLeuTtYValleuLysValSerGIuLysGI 167
518 ACAAAAGCCMAAATGSCATATGATGATGACTGTGNAATCCATGCGCAG 567
|||||
167 uGlnThraLalaLysAlaIlelTrpIleAspGysGIuIleHisAlaArg 184
568 AATGATCTCTCTGCTTCTGCTGTGCTGCTCATAGCCCAT..... 608
|||||
184 lntRpIleSerProAlaPheCysLeuTrpPheIleGIuHisIleThnGIa 200
608 ..... 608
201 pheTtYGIyleIleleGIuInlyThraSnleuLeuArgLeuValAspPh 217
608 ..... 608
217 eTtYValMetProValValAsnValAspGIuTtYAspTtYSerTrpLysL 234
609 . AATGCAATGTGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTCG 656
|||||
234 yasnArGhetTrpArgLysasnArgSerPheTtYtAlaasnHisCys 250
657 ATCGAAGACAGCTGAATAGCACTTTGTCTCCAACACAGCTGCTGAGGA 706
|||||
251 lIleGIuThraSpleuasnArgasnPheAlaSerLysHisTrpCysGIuGI 267
707 AGGTGATCCAGTTCTCATGCTCGGAAACCTACTGTGCACTTATCTTG 756
|||||
267 uGIuAlaSerSerSerSerCysSerGIuThryCysGIuLysLeuTyrProG 284
757 AGTCAGAACCAGAAAGTCAGGCGTACGTAGTTCTTGAGAAAGAATATC 806
|||||
284 luserGIuProGIuValLysAlaValAlaSerPheLeuArgAsnIle 300
807 AACACAGATTAAGCATACATACATGATGATCTTACTCCAGCATATAGT 856
|||||
301 asnGlnIleLysAlaLysIleSerMetHisSerTtYSerGIuHisIleVa 317
857 GTTTCATATTCCTATACAGAAATAAAAGCAAAAGCATGAGAACTGT 906
|||||
317 lPheProTtYSerTtYThraArgSerLysSerLysAspHisGIuGIuLys 334
907 CTCTAGTACGACGAGCAAGTTCGTGCTATTGACAAAAGTAAATAAT 956
|||||
334 eTleuValAlaSerGIuAlaValAlaArgAlaIleGIuLysThraSerLysAsn 350
957 ACCAGGTATACACATGCGCATGGCTCAGAAACCTTATACCTACTGCTG 1006
|||||

```

```

351 ThrArgTtYThraHisGIuHisGIySerGIuThraLeuTtYLeuAlaProGI 367
1007 AGGTGGGAGCATGTGATGATGATTTGGGCATCAATATTCGTT..... 1051
|||||
367 yGIuGIuAspAspTrpIleTtYtAlaPheAlaValAlaValSerLysIleAlat 384
1052 .....TAC 1054
384 leGIuLeuArgAspThraGlyThraTtYtGlyPheLeuLeuProGIuArgTtY 400
1055 ATCAACCCGACGCTGAGAGACGTTTGGCGTGTCTCTCAATATGCTTG 1104
|||||
401 lIleLysProThraCysArgGIuAlaPheAlaValAlaValSerLysIleAlat 417
1105 GCATGTCATTAGCAATGTT 1123
|||||
417 pHisValIleArgAsnVal 423

```

seq_name: /cgn2.6/ptodata/1/1aa/5A.COMB.pep:US-08-277-540-3

seq_documentation_block:

; Sequence 3, Application us/08277540

; Patent No. 5474901

; GENERAL INFORMATION:

; APPLICANT: Drayna, Dennis T., Eaton, Dan L.

; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: patln (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/277,540

; FILING DATE: 19-JUL-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/167727

; FILING DATE: 15-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/959944

; FILING DATE: 14-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/649591

; FILING DATE: 01-FEB-91

; ATTORNEY/AGENT INFORMATION:

; NAME: Hasak, Janet E.

; REGISTRATION NUMBER: 28,616

; REFERENCE/DOCKET NUMBER: 689D1C1D1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-1896

; TELEFAX: 415/952-9881

; TELETYPE: 910/371-7168

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 423 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; US-08-277-540-3

alignment_scores:

Quality: 1885.00

Ratio: 5.122

Percent Similarity: 86.998

Length: 423

Gaps: 2

Percent Identity: 86.525

NAME: Hasak, Janet E.
 REGISTRATION NUMBER: 28,616
 REFERENCE/DOCKET NUMBER: 689D1C1D1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-1896
 TELEFAX: 415/952-9881
 TELE: 910/371-7168
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 423 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-430-787A-3

alignment_scores:

Quality: 1885.00 Length: 423
 Ratio: 5.122 Gaps: 2
 Percent Similarity: 86.998 Percent Identity: 86.525

alignment_block:

US-09-980-881-1 x US-08-430-787A-3

Align seg 1/1 to: US-08-430-787A-3 from: 1 to: 423

```

18 ATGAAGCTTTGACGCTTCGATCCATTCCTTCCTGGA 67
   |||
1  MetLysLeuYsSerLeuAlaValLeuValProIleValLeuPheCysG 17
68 GCAGCATGCTTGGCGTTCGAGAGTGGCAAGTTCTAGCTGCTTCCTTA 117
   |||
17 uGlnHsValPheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuPro 34
118 GAACCTCTAGGCAAGTTCAGATCTACAGATCTTACTACACATATAG 167
   |||
34 rghrSeratrglnValGlnValLeuGlnAsnLeuThrThrThrTyGlu 50
168 ATGTGTCCTCGGACCGCGTAACAGCTGACCTTATTGTGAAGAAAAA 217
   |||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG 67
218 AGTCCATTTTGTAAATGCATCTGATGTCGACAAATGGAAGCCCAT 267
   |||
67 nValHsPhePheValAsnAlaSerAspValAspAsnValLysAlaHs 84
268 TAAATGTGAGCGGAATTCATGCAGTGTCTTGGCGAGACGTGAAGAT 317
   |||
84 euAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
318 CTATATCAACAGAGATTTCGAACGACAGCTAGCCCCGACCTCCGC 367
   |||
101 LeuIleGlnGlnIleSerAsnAspThrValSerProAlaSerAl 117
368 ATCGTACTATGACAGTATCAGTCACTCAATGAATCTATTCCTGATG 417
   |||
117 aSerTyTyGlnGlnTyThrHsSerLeuAsnGlnIleTySerTrpIle 134
418 AATTATTAAGTGAAGGATCTGATATGCTTACAAAAATCCACATTGGA 467
   |||
134 lUpheIleThrGlnuArgHsProAspMetLeuThrLysIleHsIleGly 150
468 TCCTCATTTGAGAGTACCCACCTCATGTTTAAAGTTCTTGGAAAAA 517
   |||
151 SerSerPheGlnuTyTyProLeuTyValLeuLysValSerGlyLys 167
518 ACAAAACAGCCAAAATGCATATGATGATGATGATGATGATGATGATG 567
   |||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHsAlaArg 184
568 AATGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
   |||
184 lUTrpIleSerProAlaPheCysLeuTrpPheIleGlyHsIleThrGln 200

```

```

608 ..... 608
201 PheTyrglyIleIleGlyGlnTyThrAsnLeuLeuArgLeuValAsp 217
608 ..... 608
217 eTyrValMetProValValAsnValAspGlyTyArgSpTySerTrpLys 234
609 ..AATGAAATGTGGAGAAAGACCGTCTTCTTGGGAACATCATTCG 656
   |||
234 yAsnAlaGMetTrpArgLysAsnArgSerPheTyAlaAsnHsHsCys 250
657 ATCGGAACAGACCTGAATAGCACTTGTCTCCAAACAGCTGTGAGAGA 706
   |||
251 IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHsIleTrpCysGln 267
707 AGTGCATCCAGTTCTCATGCTCGGAACCTACTGTGACTTATTCCTG 756
   |||
267 uGlyAlaSerSerSerSerCysSerGlnTrpTyCysGlyLeuTyProG 284
757 AGTCAGAACAGAAAGTGAAGGAGTGGCTAGTTCTTGAGAGAAATATC 806
   |||
284 lUserGlnProGlnValLysAlaValAlaSerPheLeuArgArgAsnIle 300
807 AACCGATTAAAGCATCATCAGCATCATTCATCTCCAGCATATAGT 856
   |||
301 AsnGlnIleLysAlaTyrlleSerMetHsSerTySerGlnHsIleVal 317
857 GTTCCATATTCCTATACAGAAATAAACAAAGACATGAGAACTGT 906
   |||
317 lPheProTySerTyThrArgSerLysSerLysAspHsGlnGluLeuS 334
907 CTCTAGTAGCAGTGAAGAGCTTCGTCTATGTGACAAATAGTAAATAT 956
   |||
334 eLeuValAlaSerGlnAlaValAlaArgAlaIleGluLysThrSerLysAsn 350
957 ACCAGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTG 1006
   |||
351 ThrArgTyThrHsGlyHsGlySerGlnThrLeuTyLeuAlaProG 367
1007 AGTGGGAGCATTTGATCTATGATTTGGCATCAATATTCCTT ..... 1051
   |||
367 yGlyGlyAspAspTrpIleTyArgPheGlnGlyIleLysTySerPheThr 384
1052 .....TAC 1054
384 lGluLeuAspArgPheGlnTyThrTyGlyPheLeuLeuProGlnuArgTy 400
1055 ATCAAAACCAACCTGTAGAGAGCTTTGCCGTGCTCTTAAATAGCTTG 1104
   |||
401 lLeuLysProThrCysArgGlnAlaPheAlaValSerLysIleAlaTr 417
1105 GCATGTCAATTAGGAATGTT 1123
417 PHisValIleArgAsnVal 423

```

seq.name: /cgn2_6/plodata/1/1aa/5B_COMB.pep:US-08-869-057-2

seq documentation_block:

Sequence 2, Application US/08869057

Patent No. 5985562

GENERAL INFORMATION:

APPLICANT: Morser, Michael J

APPLICANT: Nagashima, Mariko

TITLE OF INVENTION: Method of Detecting Thrombotic Disease

TITLE OF INVENTION: Risk

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Berlex Biosciences Legal Department

STREET: 15049 San Pablo Avenue

CITY: Richmond

STATE: California

COUNTRY: USA

```

: ZIP: 94804-0099
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/869,057
: FILING DATE: 03-JUN-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Washien, Wendy L.
: REGISTRATION NUMBER: 36,301
: REFERENCE/DOCKET NUMBER: 51509AUSM1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 510-262-5411
: TELEFAX: 510-262-7095
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 423 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: TISSUE TYPE: Plasma
: FEATURE:
: NAME/KEY: Peptide
: LOCATION: 23..401
: US-08-869-057-2

alignment_scores:
  Quality: 1885.00      Length: 423
  Ratio: 5.122          Gaps: 2
  Percent Similarity: 86.998      Percent Identity: 86.525

alignment_block:
  US-09-980-881-1 x US-08-869-057-2 ..

Align seg 1/1 to: US-08-869-057-2 from: 1 to: 423

18 ATGAGCTTGGACGCTTGCGATGCTTGTACCATTTGTTCTCTCTGTA 67
   |||||||
1  MetLysLeuYsSerLeuAlaValLeuValProIleValLeuPheCysGI 17
68 GCAGCATGTCTTGGCGTTCCAGAGTGGCGAAGTTCTAGCTGCTCTGCTA 117
   |||||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaValLeuProA 34
118 GAACCTCTAGCAAGTTCAGATTCTACAGAACTTACTACAACTATGAG 167
   |||||||
34 rghrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThyGlu 50
168 ATGTGTTCTGGCAGCGGTTACAGCTGACCTTATGTGAGAAAAAACA 217
   |||||||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGI 67
218 AGTCATTTTGTAAATGATCATGTGATGCGACAATGTGAAGCCCAT 267
   |||||||
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
268 TAAATGTAGCAGAAATTCATGCAAGTGTCTTGTGGCAGACGTGGAAGAT 317
   |||||||
84 euAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
318 CTATTATCAACAGAGATTCCACAGACAGCTAGCCCCGAGGCTTCGC 367
   |||||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
368 ATGTACTATGAACAGATATCAGTACATCAATAATGAATATATCTTGAGTAG 417
   |||||||
117 aserTyTrpGluGlnTrpHisSerLeuAsnGluIleTrpSerTrpIleG 134

```

```

418 AATTATTAAGTGAAGGCAATCTGATATGCTTACAAAAATCCATTTGA 467
   |||||||
134 IupheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
468 TCCATATTGGAGAGTACCCCACTATGTTTAAAGTTTGTGAAANA 517
   |||||||
151 SerSerPheGluLysTyTrpLeuTrpValLeuLysValSerGlyLysGI 167
518 ACAACAGCCCAAAATGCAATGATGATGAGTGTGGAATCCATGCCAGAG 567
   |||||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
568 AATGATCTCTCCTGCTTCTGCTGTGTTCAATAGCCAT..... 608
   |||||||
184 IuTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
608 ..... 608
201 PheTyrglyIleIleGlyGlnTrpThrAsnLeuLeuArgLeuValAspPh 217
608 ..... 608
217 eTyValMetProValValAsnValAspGlyTyAspTrpSerTrpLysL 234
609 ..AATCGAATGTGAGAAAGAACCGTTCTTCTATGCGAACAATCATTC 656
   |||||||
234 yAsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnAsnHisCys 250
657 ATCGGAACAGACCTGCAATAGCACTTGTCTCCAAACAGCTGTGTGAGA 706
   |||||||
251 IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGI 267
707 AGTGCATCCAGTTCCTCATGCTCGGAACCTAGTGGCAATTATGCTCG 756
   |||||||
267 uGlyAlaSerSerSerSerCysSerGluThrTrpCysGlyLeuTrpArg 284
757 AGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTGAGAAATAATATC 806
   |||||||
284 IuSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 300
807 AACCAATTAAAGCATACATGACATGCAATTCATCTCCAGCATATAGT 856
   |||||||
301 AsnGlnIleLysAlaTyTrpIleSerMetHisSerTySerGlnHisIleVal 317
857 GTTCCAAATTCCTATACAGAGTAAAGCAAGCAACCATGAGGAACGT 906
   |||||||
317 IPherProLysSerTyThrArgSerLysSerLysAspHisGluGluLeuS 334
907 CTCTAGTAGCCAGTGAAGCAGATGCTGCTATTGACAAATAGTAAAT 956
   |||||||
334 erLeuValAlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsn 350
957 ACCAGGTATACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCGG 1006
   |||||||
351 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyTrpLeuAlaProGI 367
1007 AGTGGGAGCATGGAATCTATGATTTGGGCATGCAATATTCGTT..... 1051
   |||||||
367 yGlyGlyAspAspTrpIleTyAspLeuGlyIleLysTySerPheThrI 384
1052 .....TAC 1054
384 legIuLeuArgAspThrGlyThrTyrglyPheLeuLeuProGluArgTy 400
1055 ATCAAAACCACTGTAGAGAAAGCTTTGCCGCTGTCTTAAATAAGCTTG 1104
   |||||||
401 IleLysProThrCysArgGluAlaPheAlaValAlaSerLysIleAlaTr 417
1105 GCATGTCAATTAGGAATGTT 1123
   |||||||
417 PhisValIleArgAsnVal 423

```

```
seq_name: /cgn2_6/plodata/1/1aa/5A_COMB.pep:us-07-649-591B-7
seq_documentation_block:
; Sequence 7, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
US-07-649-591B-7

alignment_scores:
Quality: 655.00 Length: 418
Ratio: 2.472 Gaps: 9
Percent Similarity: 63.397 Percent Identity: 35.885

alignment_block:
US-09-980-881-1 x US-07-649-591B-7 ..
Align seg 1/1 to: US-07-649-591B-7 from: 1 to: 417

33 CTTCGAGTCCCTGTACCCATGTTCTCTGTGTGAGCAGCATGTCTGCG 82
:::|||||::: ::::: |||
7 MetAlaValIleTyrThrIleuAlaIleAlaProValHis..... 20
83 GTTCCAGAGTGGCCAAAGTTTACGTGCTCTTCTGAGAACCTCTAGGCAAG 132
:::|||||::: ::::: |||
21 PheAspArgGluValPheArgValIleuGlnAsnGluLysHisA 37
133 TTCAAGTTCTACAGATCTTACTACAGACATATGAGATTTCTCTGCGGAG 182
:::|||||:::|||||::: ::::: |||
37 IAsnValIleuLysAsnIleuThrGlnSerIleGluIleuAspHetrPyr 53
183 CCGGTACAGAGTGCACCTTATTTGTGAAGAAACAAAGTCCATTTTTGTGT 232
||| ::::: |||::: ||| ||| |||
54 ProAspAlaIleHisAspIleAlaValAsnMetThrValAspPheArgVa 70
233 AAATGCATCTGATGTGCAACATGTGAAGCCCATTTAAATGTGACGGGAA 282
::: ::::: ::::: |||:::
70 IserGluLysGlnSerGlnThrIleGlnSerThrLeuGlnHisLysI 87
```

```
283 TTCCATGCAAGTCTCTGCTGCGAGACGTGGAAGATCTTATTCACAGCAG 332
|| :::::|||||::: |||:::|||||::: |||:::|||||
87 IeHisTyrGluIleIleuIleHisAspLeuGlnGluIleGluLysGln 103
333 ATTTCG...AAGCAGACAGTCAGCCCCGAGCCTCGCAGTCAGTATGCA 379
:::|||||::: ::::: ||| ||| |||
104 PheAspValIysAspGluIleAlaGlyArgHisSer.....TyrAl 117
380 ACAGTATCACTACTAATAGATCTATCTGTGATAGATTTATATCACTG 429
:::|||||::: ::::: ||| ||| ||| |||
117 AluSerTyrAsnAspIrrAspLysIleValSerThrPheGluLysMetLeu 134
430 AGAGCATCTGATATGCTTACAAAAATCCAAATGGATCCGATTTGAG 479
|||:::|||||:::|||||::: ||| ||| ||| ||| |||
134 IuLysHisProGluMetValSerArgIleLysIleGlySerThrValGlu 150
480 AAGTACCATCTATATGTTTAAAGTTTCTGGAAGAACAAACAGCCAA 529
|||:::|||||::: |||::: |||
151 AspAsnProLeuTyrValIleuLysIle...GlyLysLysAspGlyLys 166
530 AAATGCATATGATGATGCTGTGGAATCCATGCCAGAGATGGATGCTGC 579
:::|||||:::|||||::: ||| ||| ||| ||| ||| |||
166 GluSerAlaIlePheMetAspCysGlyIleHisAlaArgGluTrpIleSer 183
580 CTGCTTCTGCTGTGGTTCATA.....GGCCAT 608
|||:::|||||::: |||::: |||
183 roAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGluLys 199
609 AAT..... 611
200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuP 216
612 .....CGAATG 618
216 oValPheAsnValAspGlyTyrIleTrpSerTrpThrGlnAspArgMet 233
619 GGAGAAAGAACCGTTTCTTATGCGAACATTCATCGAGACAGAC 668
|||:::|||||::: |||::: ||| ||| ||| ||| |||
233 rParGlyAsnArgSerArgAsnGlnAsnSerThrCysIleGlyThrAsp 249
669 CTGATAGCAACCTTGTCCCAACACGTGTGAGAGAGGTGATCCAG 718
|||:::|||||::: |||::: |||
250 LeuAsnArgAsnPhe...AspValSerTrpAspSerSerProAsnThr 265
719 TTCCATGCTGCGAACCTACTGTGACTTTATCTCGAGTGCAGACAG 768
::: ||| ::::: ||| ||| ||| ||| ||| |||
265 nLysProCysLeuAsnValTyrArgGlyProAlaProGluSerGluLys 282
769 AAGTGAAGCAGTGTGCTTCTTGTGAGAAATATCAACAGATTTAA 818
|||:::|||||::: |||::: ||| ||| ||| ||| |||
282 IuThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
819 GCATACATGAGCATGATCATCTACCTCCAGCATATATGTTTCCATATTC 868
|||:::|||||::: |||::: ||| ||| ||| ||| |||
299 AlatYrIleThrPheHisSerTyrSerGlnMetLeuIleProTyrGln 315
869 CTATACACAGAGTAAAGCAAGACCATGAGAACTGTCTAGTAGGCA 918
||| ||| ::::: |||::: ||| ||| |||
315 YTrGlyThrPheLysLeuProProAsnHisGlnAspLeuLysValAlaA 332
919 GTGAAGCAGTGTGCTATTTGACAAACTAGTAAATATCAACAGATTTAC 968
|||::: |||::: |||::: ||| ||| ||| ||| |||
332 rGluAlaThrAspAlaLeu...SerThrArgTyrGluThrArgTyrIle 347
969 CATGGCAATGGCTCAGAAACCTTATACCAAGTCTCGAGAGTGGGAGCA 1018
:::|||||::: ::::: |||::: ||| ||| ||| |||
348 TyrGlyProIleAlaSerThrIleTyrLysThrSerGlySerSerLeuAs 364
1019 TTGAGTATATGATTTGGCATCAATATATCGTTTAC..... 1054
|||:::|||||::: |||::: ||| ||| ||| ||| ||| |||
364 rTrpValTyrAspLeuGlyIleLysHisThrPheAlaPheGluLeuAlaG 381
```



```

299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrG1 315
869 CTATACAGCAAGTAAAGCAAGACCATGAGAACTGTCTTAGTACCA 918
      |||||      :|||:|||||      |||||
315 yTyrThrPheIleuProGlnHisGlnAspLeuLeuValAla 332
      |||||      :|||:|||||      |||||
919 GTGACGACGTCTGCTATTGACAAAAGTGTAAATACCAGGTATACA 968
      |||||      :|||:|||||      |||||
332 rGlleIleThrAspAlaLeu...SerThrArgTyrGlnThrArgTyrIle 347
      |||||      :|||:|||||      |||||
969 CATGCCATGCTGACAAACCTTATACCTGATCTGAGAGGGAGCA 1018
      :|||:|||||      :|||:|||||      :|||:|||||
348 TyrGlyProIleAlaSerThrIleTyrLysThrSerGlySerSerLeu 364
      :|||:|||||      :|||:|||||      :|||:|||||
1019 TTGGATCTATGATTTGGGACATCAATATTCGTTAC..... 1054
      |||||      :|||:|||||      :|||:|||||
364 pTrpValTyrAspLeuGlyIleLysHisThrPheAlaPheGlnLeuArg 381
      |||||      :|||:|||||      :|||:|||||
1055 .....ATCAACCCACC 1066
      |||||      :|||:|||||      :|||:|||||
381 sPrlsGlyLysSerGlyPheLeuLeuProGlnSerArgIleLysProThr 397
      |||||      :|||:|||||      :|||:|||||
1067 TGTAGAGAAAGCTTTGCCGCTGTCTTAAATATAGCTTGGCATGCAATAG 1116
      |||||      :|||:|||||      :|||:|||||
398 CysLysGlnThrMetLeuSerValLysPheIleAlaLysTyrIleLeu 414
      |||||      :|||:|||||      :|||:|||||
1117 GAAT 1120
      |||
414 sAsn 415

```

seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-430-787A-7

```

seq_documentation_block:
: Sequence 7, Application US/08430787A
: Patent No. 5593674
: GENERAL INFORMATION:
: APPLICANT: Drayna, Dennis T., Eaton, Dan L.
: TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/430.787A
: FILING DATE: 27-APR-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/277,540
: FILING DATE: 19-JUL-1994
: APPLICATION NUMBER: 08/167727
: FILING DATE: 15-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/959944
: FILING DATE: 14-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/649591
: FILING DATE: 01-FEB-91
: ATTORNEY/AGENT INFORMATION:
: NAME: Haack, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 689D1C1D1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-1896
: TELEFAX: 415/952-9881

```

```

: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 417 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: US-08-430-787A-7

alignment_scores:
      Quality: 655.00      Length: 418
      Ratio: 2.472      Gaps: 9
      Percent Similarity: 63.397      Percent Identity: 35.885

alignment_block:
US-09-980-881-1 x US-08-430-787A-7 ..
Align seq 1/1 to: US-08-430-787A-7 from: 1 to: 417

33 CTTCAGTCCCTGTGACCCATGTTCTCTCTGTGAGCAGCATGCTCTGCG 82
      :|||:|||||      :|||:|||||      :|||:|||||
7 MetAlaValIleTyrThrThrIleAlaIleAlaProValHis..... 20
      :|||:|||||      :|||:|||||      :|||:|||||
83 GTTCCAGAGTGGCCAGTCTTACGCTGCTTCTTACAGCCTTACAGCAG 132
      :|||:|||||      :|||:|||||      :|||:|||||
21 .PheAspArgGlnLysValAlaPheArgValLysLeuGlnAsnGlnLysHis 37
      :|||:|||||      :|||:|||||      :|||:|||||
133 TTCAAGTTCTACAGAACTTACTATACAAACATATGAGATGTTCTCTGCG 182
      :|||:|||||      :|||:|||||      :|||:|||||
37 IAserValIleuLysAsnLeuThrGlnSerIleGlnLysLeuAspPheThr 53
      :|||:|||||      :|||:|||||      :|||:|||||
183 CCGGTAAACAGCTGACCTTATTTGTAAGAAAAAACAAGTCCATTTTGT 232
      :|||:|||||      :|||:|||||      :|||:|||||
54 ProAspAlaIleHisAspIleAlaValAlaMetThrValAspPheArg 70
      :|||:|||||      :|||:|||||      :|||:|||||
233 AAATGCATCGATGTCGACAAATGTAAGCCCATTTAAATGTACGGGAA 282
      :|||:|||||      :|||:|||||      :|||:|||||
70 IserGlnLysGlnSerGlnThrIleGlnSerThrLeuGlnHisLysI 87
      :|||:|||||      :|||:|||||      :|||:|||||
283 TTCCATGACAGTGTCTGCTGCGACAGTGGAGATGTTATTAACAGCAG 332
      :|||:|||||      :|||:|||||      :|||:|||||
87 LeHisTyrGlnIleLeuIleHisAspLeuGlnGlnGlnLysGln 103
      :|||:|||||      :|||:|||||      :|||:|||||
333 ATTTC...AACGACAGTCAAGCCCCGAGCCTCCGATCGTACATGA 379
      :|||:|||||      :|||:|||||      :|||:|||||
104 PheAspValLysAspGlnIleAlaGlyArgHisSer.....TyrAl 117
      :|||:|||||      :|||:|||||      :|||:|||||
380 ACGATATCACTCAATGAATGATATGTTGATAGAAATTTATACATG 429
      :|||:|||||      :|||:|||||      :|||:|||||
117 AluSTyrAsnAspTrpAspLysIleValSerThrGlnLysMetLeuG 134
      :|||:|||||      :|||:|||||      :|||:|||||
430 AGAGCATCCTGATATGCTTACAAAATCCAAATGATGATCCATATTGAG 479
      :|||:|||||      :|||:|||||      :|||:|||||
134 IuLysHisProGlnMetValSerArgIleLysIleLysThrValGln 150
      :|||:|||||      :|||:|||||      :|||:|||||
480 AAGTACCCACTATATGTTTAAAGTTTCTGAAAAGAACAAACAGCCAA 529
      :|||:|||||      :|||:|||||      :|||:|||||
151 AspAsnProLeuTyrValLeuLysIle...GlyLysLysAspLysLys 166
      :|||:|||||      :|||:|||||      :|||:|||||
530 AAATGCCATATGATGATGCTGCGATCCATGCCAAGAAATGATGTCG 579
      :|||:|||||      :|||:|||||      :|||:|||||
166 GlyLysAlaIlePheMetAspCysGlyIleHisAlaArgLysTrpLeu 183
      :|||:|||||      :|||:|||||      :|||:|||||
580 CTGCTTCTGCTTGGTGCAT.....GGCAT 608
      :|||:|||||      :|||:|||||      :|||:|||||
183 roAlaPheCysGlnThrPheValTyrGlnAlaThrLysSerTyrGlyLys 199
      :|||:|||||      :|||:|||||      :|||:|||||
609 AAT..... 611
      :|||:|||||      :|||:|||||      :|||:|||||
200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuPr 216
      :|||:|||||      :|||:|||||      :|||:|||||
612 .....CGAATGT 618
      :|||:|||||      :|||:|||||      :|||:|||||

```

```

216 ovalpheasnvalaspolytyriletpsertrphtglnaspargmect 233
619 GGAAGAAAGACGCTTCTTATGCGAACAATCATTCATCGAAGAC 668
233 rpatgylasnasrargserargasnlnasnsertlrcyslleglythrasp 249
669 CTGAATGCACTTGTCTCCAAACACTGCTGTAGAGAGAGTCATCCAG 718
250 leuasnargasnph...Aspvalsetrpaspserseproasntrhas 265
719 TTCCTCATGCGGGAACCTACTGTGACCTTATCCGAGCAGACACAG 768
265 nlvsprocysleuasnvaltyrargglyproalaproglinserglulysg 282
769 AAGTAAGCAGAGTGGTAGTCTTGTAGAGAAATATCAACAGATTAA 818
282 luthrlysalalvalthrashnphleargserhslsleasnserileys 298
819 GCATACATGACATGATCTACTCCAGCATATGCTGTCCATATTC 868
299 Alatyrllethrphnhsertysertglmetleuileuprotygl 315
869 CTATACAGAGTAAAGCAAGACATGAGAGACTGTCTAGTACCCA 918
315 YTYTthrphelysleupropoasnhslnaspleuulysvalalaa 332
919 GTGAGCAGTCTGTCTATTGCAAACTAGTAAATATCCAGATATACA 968
332 rgllealathraspalaleu...Serthrargtyrgluthrargtyrle 347
969 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGAGCA 1018
348 Tyrclyprolealaserthriletyrlysrserglyserseleuasn 364
1019 TTGATCATGATTTGGGCATCAAAATATGCTTAC..... 1054
364 prtrvaltyrpsleuglylelyshsthrphnlealphegluleuarg 381
1055 .....ATCAAAACCCACC 1066
381 splysglylyserglypheleuleuprogulserarglileysprothr 397
1067 TGTAGAGACGTTTGGCCGTGTCTTAAATAGCTTGCGCATGTCATTAG 1116
398 Cyslysglutrhmetleuservallysheillealalystryrileu 414
1117 GAAT 1120
414 SAsn 415

```

seq_name: /cgn2_6/prodata/1/laa/5B_COMB.pep:US-08-860-882A-57

seq_documentation block:

```

; Sequence 57. Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGEITE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUM
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.

```

```

; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-860-882A-57

```

alignment_scores:

```

Quality: 651.50 Length: 419
Ratio: 2.404 Gaps: 8
Percent Similarity: 64.678 Percent Identity: 33.890

```

alignment_block:

US-09-980-881-1 x US-08-860-882A-57 ..

Align seg 1/1 to: US-08-860-882A-57 from: 1 to: 415

```

33 CTGCGAGTCTGTACCCATTGCTCTGTGAGCAGATGCTGCG 82
2 leuulvalleuvalthrvalalaleuallaserhslshslsglygl 18
83 G...TTCCAGAGTGGCCAGTTCTAGCTGCTCTTGAACCTGTAGCC 129
18 whishegluglyglulysvalrphnargvalasnvalglulaspgrluasn 35
130 AAGTCAGTCTACAGAACTTACTACACATATGAGATTGTCTGG 179
35 isileasnillearggluleualaserthrthrghlnleasprhetrp 51
180 CAGCCGTAACAGCTGACCTTATGTGAAGAAACAAAGTCCATTGTT 229
52 lvsproaspservalthrghlnlelysprohlsertlhrvalasprhar 68
230 TGTAAATGATCTGATGTGACATATGGAAGCCATTAAATGTGAGCG 279
68 gvallysalagluasprthrvalthrvalglulasnvalleuysglnasn 85
280 GAATTCATGACAGTGTCTGTGCGACAGCGTAGAGATCTTATCAAG 329
85 lulnleghltyrlysalvalleuileaserasnleuargasnvalalglula 101
330 CAGATTCCACAGCAGACGACGAGCCCGAGCCTCGCATCTACTATGA 379
102 glnphnaspserargval.....Argalathrghlnhserttygl 115
380 ACAGTATCACTCAGTAATGAATGTATTTGTGATPACATTTAATAC 429
115 ulstytrasnlystrpghluthrilleglualatprhnglnvalalat 132
430 AGAGCATCCGATATGCTTACAAAATCCACATTCAGTCCATTTGAG 479
132 hrghlnasprloalaleuileaserargservalileglythrthrphn 148

```

```

480 AAGTACCACTCTATGTTTAAAGTTTCGAAAACAAACACCA 529
      ::::::::::::::::::::
149 GYAAGTAAATTAATTAATTAATTAATTAATTAATTAATTA 164
      ::::::::::::::::::::
530 AATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 579
      ::::::::::::::::::::
164 AATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 181
      ::::::::::::::::::::
580 CTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
      ::::::::::::::::::::
181 TGAATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 197
      ::::::::::::::::::::
602 .....
198 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 214
      ::::::::::::::::::::
603 .....
214 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 231
      ::::::::::::::::::::
619 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 668
      ::::::::::::::::::::
231 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 247
      ::::::::::::::::::::
669 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 718
      ::::::::::::::::::::
248 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 263
      ::::::::::::::::::::
719 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 768
      ::::::::::::::::::::
263 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 280
      ::::::::::::::::::::
769 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 818
      ::::::::::::::::::::
280 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 296
      ::::::::::::::::::::
819 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 868
      ::::::::::::::::::::
297 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 313
      ::::::::::::::::::::
869 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 918
      ::::::::::::::::::::
313 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 330
      ::::::::::::::::::::
919 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 968
      ::::::::::::::::::::
330 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 345
      ::::::::::::::::::::
969 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1018
      ::::::::::::::::::::
346 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 362
      ::::::::::::::::::::
1019 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1057
      ::::::::::::::::::::
362 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 379
      ::::::::::::::::::::
1058 .....
379 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 395
      ::::::::::::::::::::
1067 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1116
      ::::::::::::::::::::
396 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 412
      ::::::::::::::::::::
1117 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT
      ::::::::::::::::::::
412 GATGAGCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT
      ::::::::::::::::::::
seq_name: /cgn2_6/ptodata/1/lae/5A-COMB.pep:US-07-649-591B-6
seq_documentation_block:
: Sequence 6, Application US/07649591B

```

```

: Patent No. 5206161
: GENERAL INFORMATION:
: APPLICANT: Dennis Drayna and Daniel Eaton
: TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/649,591B
: FILING DATE: 19910201
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasak, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 689
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/266-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 417 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: US-07-649-591B-6

alignment_scores:
      Quality: 647.50      Length: 427
      Ratio: 2.398        Gaps: 10
      Percent Similarity: 63.232      Percent Identity: 33.489

alignment_block:
US-09-980-881-1 x US-07-649-591B-6 ..
Align seg 1/1 to: US-07-649-591B-6 from: 1 to: 417
6 ATGCTGTTGGATGAAGCTTTGCGAGCTTGCAGCTGCTGTTACCATTTGT 55
  ::::::::::::::::::::
5 LeuProValGlyLeuLeuAlaThrThrLeuAlaIle...AlaProVal.. 19
56 TCCTCTTGTGACACAGATGCTTCCGCTTCCAGAGTGGCAAGTCTGAG 105
  ::::::::::::::::::::
20 .....
106 CTGCTCTTCCATTAACCTTACGAGCTTGCAGTTCAGTTCATGAGTATCTTACT 155
  ::::::::::::::::::::
28 rGValysProGlnAspGluLysGlnAlaAspIleIleLysAspLeuAla 44
156 ACAACATATGAGATGTTCTGCGACGCGGTAAACAGTGCATTAATTTGT 205
  ::::::::::::::::::::
45 LysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisValAl 61
206 GAAGAAAAACAAGTCCATTTTGTAAATGATGATGATGATGATGATGATG 255
  ::::::::::::::::::::
61 AlaAsnMetMetValAspPheArGValSerGluLysGlnSerGlnAlaI 78
256 TGAAGCCCATTTAAATGAGCGGGAATTCATGCAATGCAATGCAATGCAATG 305
  ::::::::::::::::::::
78 IeGlnSerAlaLeuAspGlnAsnLysMetHisTyrGlnIleLeuIleHis 94

```



```

306 GACGTGGAAGATCTTATTCACAGCATTTCC...AACGACACAGTCAG 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 Aspleungingluileg1ulysglinpheaspyallysgluasp1leapr 111
353 CCCCCGACCCCGATGCTACTATGACATATACACTACATAAAGAA 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
111 cglAArgHHisSer.....TyrAlaLysTyrAsnAsnTrpGluLysI 125
403 TCTATTTCTGGATAGATTTATTAAGTGAAGCATCTCTGATATGCTTCA 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
125 leValAlatrPrThrgluLysMetMetAspLysTyrProgluMetValSer 141
453 AAAATCCATGATGCTCTCATTTGAGAGATGACCATCTATGTTTAA 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
142 ArgileLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLys 158
503 GGTTCCTGGAAAGAACAAACAGCCAAAATGSCATATGATGATGCTGTG 552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
158 sile...GlygluLysAsnGluArgLysAlaIleLeuMetAspCysG 174
553 GAATTCATGCCAGAGATGATGATCTCTGCTTCTCTGCTGTTGCATTA 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 LylIeHisAlaIArgLutPrValSerProAlaPheCysGlnTrpPheVal 190
602 ..... 602
191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLe 207
603 .....GGCCAT. 608
207 uAspArgMetAsnDheTyrIleLeuProValPheAsnValAspGlyTyrI 224
609 .....AATCGAATGTGGAGAAAGAACCGTTCTTTCAT 641
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 letPrSerTrpThrLysAsnArgMetTrpArgLysAsnArgSerLysAsn 240
642 GCGAACAATCTTGCATCGGACAGACCTGAATAGCAACTTGTCTCAA 691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
241 GlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe...AsnAl 256
692 ACACGTGATGAGGAGGTGCATCGATTCCTCATGCTCGCAAACTCT 741
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
256 aSerTrpAsnSerIleProAsnTrpAsnAspProCysAlaAspAsnTyrA 273
742 GTGGACTTTATCTGAGTCAGACACAGAGTAAGAGCGAGTGGCTAGTTTC 791
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
273 rglLysSerAlaPrGluSerGluLysGluThrLysAlaValThrAsnPhe 289
792 TTGAGAAAGAAATATCAACAGATTAAAGCATATCAGCATGATTCATA 841
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyr 306
842 CTCACAGATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAG 891
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
306 rSerGlnMetLeuLeuPheProTyrGlyTyrThrSerLysLeuProProA 323
892 ACCATGAGGAAGTCTCTAGTACAGAGTGAAGAGTCTGCTATTTGAC 941
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
323 snHisGluAspLeuAlaLysValAlaLysIleGlyThrAspValLeu... 338
942 AAAACTGTAATAATACAGGATATACACATGGCATGGCTCAGAAACCTT 991
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
339 SerThrArgTyrGluThrArgTyrIleGlySerProIleGluSerThrI 355
992 ATACCTAGCTCTGAGGTGGAGCATGATGATGATGATGATGATGATGAT 1041
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
355 eTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleL 372
1042 AATATTCGTTTAC..... 1054
372 yshHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

```

```

1055 .....ATCAAAACCACCTGTAGAGAACTTTTGCCGCTGT 1089
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaVa 405
1090 CTCCTAAATAGCTTGGCATGCTCATAGCAAT 1120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
405 llyspHeleAlaLysTyrIleLeuLysHis 415
seq_name: /cgn2_6/ptodata/1/lae/5A_COWB pep:US-08-277-540-6
seq_documentation block:
; Sequence 6, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Baton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Haasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-277-540-6
alignment_scores:
Quality: 647.50 Length: 427
Ratio: 2.398 Gaps: 10
Percent Similarity: 63.232 Percent Identity: 33.489
alignment_block:
US-09-980-881-1 x US-08-277-540-6 ..
Align seg 1/1 to: US-08-277-540-6 from: 1 to: 417
6 AATGCTGTGGAGATGAAGCTTTGACCGCTTGAGTCAGTCTTGACCACTGT 55
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 LeuProValGlyLeuIleAlaThrLeuAlaIle...AlaProVal.. 19
56 TCTCTCTGTAGAGCATGCTTCGCGTTCAGAGATGGCAAGTTCTAG 105

```

```

20 ..... ||||| :|||:
106 CTGCTTCCTAGAACCTTAGCAAGTTCAGTTCACAGATCTTACT 155
    ||||| :|||: |||||: |||||: |||||: |||||:
28 rGValLysPrGlnAspGlnLysGlnAlaAspIleIleLysAspLeuAla 44
    ||||| :|||: |||||: |||||: |||||: |||||:
156 ACAACATATGAGATTGTTCTCTGGACCCGGTAAACAGTCACCTATTGT 205
    ||||| :|||: |||||: |||||: |||||: |||||:
45 LysThrAsnGlnLeuAspPheTrpTyrProGluAlaThrHisValAla 61
206 GAGGAAAAACAGTCAATTTTGTAAATGCATCTGATGTCCACATG 255
    :|||: |||||: |||||: |||||: |||||: |||||:
61 AlaAspMetMetValAspPheArgValSerGlnLysGlnSerGlnAla 78
256 TGAAGCCCACTTAAATGTGACGGAATTCATGCAGTCTCTCTGCA 305
    :|||: |||||: |||||: |||||: |||||: |||||:
78 LeuInsérAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94
306 GAGCTGAGAGATCTTATTCAACAGCAGATTTC...AACGACACAGTCAG 352
    |||||: |||||: |||||: |||||: |||||: |||||:
95 AspLeuGlnGlnGluIleGlnLysGlnPheAspValLysGlnAspIlePr 111
353 CCCCCGAGCCTCGCATCTACTATGAACAGTATCACTCACTAATGAAA 402
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
111 OGlyAlaGlnHisSer.....TyrAlaLysTyrAsnAsnTrpGluLysI 125
403 TCTATTCTGTGAGATTTATTAAGTACGAGCGCATCTGATATGCTTACA 452
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 LeValAlaTrpThrGlnLysMetLeuAspLysTyrProGluMetValSer 141
453 AAATCCACATGTGATCTCATTTGAGAGATACCCACTCATGTTTTAAA 502
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
142 ArgIleLysIleGlySerThrValGlnAspAsnProLeuTyrValLeuLys 158
503 GGTTCCTGAAAAAGAACAAACAGCAAAATGCCATATGATGATGACGTG 552
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
158 sIle...GlyGlnLysAsnGlnLysArgLysAlaIlePheMetLeuPcySg 174
553 GAATTCATGCCAGAGAAATGATCTCTCTCTCTCTCTCTCTCTCTCTCT 602
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 LysIleHisAlaArgGlnTrpValSerProAlaPheCysGlnTrpPheVal 190
602 .....
191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLe 207
603 .....GGCCAT. 608
207 uAspArgMetAsnPheTyrIleLeuProValPheAsnValAspGlyTyrI 224
609 .....AATCGAATGTGAGAAAGAACCGTCTCTTCTAT 641
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 LeTrpSerTrpThrLysAsnArgMetTrpArgLysAsnArgSerLysAsn 240
642 GCGAACAATATTCATTCGGAACAGACCTGAATAGCAACTTTGTCTCCA 691
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 GlnAspSerLysCysIleGlyThrAspLeuAsnArgAsnHe...AsnAl 256
692 ACACTGGTGTGAGGAAGGTGCATCGATCTCTGCTCGAAGACCTACT 741
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
256 AserTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrA 273
742 GTGGACTTTATCTGAGTCAAGAACAGAACTGAAGCAGTGGCTAGTTTC 791
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 rGlyLysérAlaProGlnSerGlnLysGlnThrLysAlaValThrAsnPh 289
792 TTGAGAAAGAAATATCAACGAGATTAAAGCATACATCAGCATTCATCATA 841
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 IleArgSerHisLeuAsnGlnLysValTyrIleThrPheHisSerTyr 306
842 CTCCTCAGCATATAGTGTTCATATTCCTATACAGCAAGTAAAGCAAG 891
    ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

306 rSerGlnMetLeuLeuPheProTyrGlyTyrThrSerLysLeuProProA 323
892 ACATAGAGAGACTGTCTCTACTAGCACAGTGAAGAGTTCGCTATTGAC 941
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
323 snHisGlnAspLeuAlaLysValAlaLysIleGlyThrAspValLeu... 338
942 AAAACTAGTAAATAATACAGGTATACACATGGCCATGGCTCGAAGACCT 991
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
339 SerThrArgTyrGlnThrArgTyrIleTyrGlyProIleGlnSerThrI 355
992 ATACCTAGCTCTCTGAGTGGGACAGCATGTGATATGATTTGGGCATCA 1041
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 eTyrProIleSerGlySerLeuAspTrpAlaTyrAspLeuGlyIle 372
1042 AATATCGTTTAC..... 1054
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 yshIstThrPheAlaPheGlnLeuArgAspLysGlyLysPheGlyPheLeu 388
1055 .....ATCAAAACCCACCTGTAGAGACCTTTGCCGCTGT 1089
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
389 LeuProGlnSerArgIleLysProThrCysArgGlnThrMetLeuAlaVa 405
1090 CTCATAAATAGCTTGGCATGTCAATAGCAAT 1120
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
405 LysPheIleAlaLysTyrIleLeuLysHis 415
seq_name: /cgn2_6/ptodata/1/laa/5A_COMB.pep:us-08-430-787A-6
seq_documentation block:
; Sequence 6, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patln (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids

```

TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-6

alignment_scores:
Quality: 647.50 Length: 427
Ratio: 2.398 Gaps: 10
Percent Similarity: 63.232 Percent Identity: 33.489

alignment_block:

US-09-980-881-1 x US-08-430-787A-6 ..

Align seq 1/1 to: US-08-430-787A-6 from: 1 to: 417

```

6 ATTCCTGGATGAGCTTTCAGCCCTTCAGCTTCAGTCCCTTACCCATGT 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle..AlaProVal.. 19
56 TCCTCTCTGACGACGATGCTTCGCGTTCAGAGTGGCCAGTCTTAG 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20 .....ArgPheAspArgGluValPheA 28
106 CTGCTCTCCAGACCTCTAGGCAAGTTCAGTTCACAGATCTTACT 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28 rValLysProGlnAspGlnLysGlnAlaAspIleIleLysAspLeuAla 44
156 ACAATATGATGATGCTCTCTGACGCGCGGTACAGCTGACCTTAATGT 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 LysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisValAl 61
206 GAAGAAAAACAGTCATTTTGTAAATGATCTGATGTCGACAAATG 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 aAlaAsnMetValAspPheArgValSerGluSerGlnAlaI 78
256 TGAAGCCCATTTAAATGTGAGCCGATTCATGACAGTCTGTGCGCA 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78 LeuGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuHis 94
306 GACCTGGAGATCTTATACACAGCATTTTC...AAGCAGACAGTCAG 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 AspLeuGlnGluIleGluLysGlnPheAspValLysGlnAspIlePr 111
353 CCCCCGAGCTCCGATCTGATGAAACATGATCTACTACTAAATGAA 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 oGlyArgHisser.....TyrAlaLysTyrAsnAsnTrpGluLysI 125
403 TCCTATCTGATAGATTTATTAAGTACAGGACATCCGATATGCTTACA 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 LeuAlaIleTrpThrGluLysMetMetAspLysTyrProGluMetValSer 141
453 AAATTCACATTTGATTCCTATTGAGAGTACACCATCTATGTTTAAA 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 ArgIleLysIleGlySerThrValGlnAspAsnProLeuTyrValLeuL 158
503 GGTTCCTGGAAGAACAACAGCCAAAATTCATATGATGATGACTGTG 552
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
158 sile...GlyGluLysAsnGlnArgGlyAlaIlePheMetAspCysG 174
553 GAATTCATGCCAGAGATGATCTCTGCTTCTGCTTCTGCTGTTGTTCA 602
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174 LysIleHisAlaArgLysIleValSerProAlaPheCysGlnTrpPheVal 190
602 ..... 602
191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeu 207
603 .....GGCCAT. 608
207 uAspArgMetAsnPheTyrIleLeuProValPheAsnValAspGlyTyrI 224
609 .....AATGCAATGTGGAGAAAGAACGTTCTTCAAT 641
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

224 LeuPheSerTrpThrLysAsnArgMetTrpArgLysAsnArgSerLysAsn 240
642 GCGAACATCATTTGATCGAGCAACAGACCTGATAGCACTTGTCTCCA 691
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 GlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe...AsnAl 256
692 ACACTGTGTGAGGAGAGTGCATCCAGTTCCTATGCTGGAAACCTACT 741
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
256 AserTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrA 273
742 GTGACCTTATCTGAGTCAGACAGACAGAGTGAAGCGCTGAGTTC 791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
273 rGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289
792 TTGAGAGAAATATCACACAGATTAAGCATATACATGATCATTCATA 841
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
290 IleArgSerHisLeuAsnGlnIleLysValLysIleThrPheHisSerT 306
842 CTCCGACATATAGTGTTCATATTCCTATPACAGCAAGTAAAGCAAA 891
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306 rSerGlnMetLeuLeuPheProTyrGlyTyrThrSerLysLeuProProA 323
892 ACCATGAGAGACTGCTCTAGTACGACAGTGAAGAGTTCGTATATAC 941
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323 snHisGlnAspLeuAlaLysValAlaLysIleGlyThrAspValLeu... 338
942 AAACTAGTAAATAATCACAGATATACATGACGCTGCTGAGAAACCTT 991
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThrI 355
992 ATACCTAGCTCTCGAGGAGGAGCATGATGATGATGATGATGATGATG 1041
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 eTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIle 372
1042 AATATGCTTAC..... 1054
372 yHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388
1055 .....ATCAACCCACCTGTAGAGAACCTTTGGCGCTGT 1089
389 LeuProGluSerArgGlyLeuProThrCysArgGlyThrMetLeuAlaVa 405
1090 CTCTAAATAGCTTGGCATGTCATTAGAAT 1120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 LysPheIleAlaLysTyrIleLeuLysHis 415
seq_name: /cgn2_6/protodata/1/laa/5A_COMB.pep:US-08-696-139-2
seq_documentation_block:
; Sequence 2, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hersberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Steiner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```



```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-125

alignment_scores:
    Quality: 635.50      Length: 422
    Ratio: 2.345         Gaps: 8
    Percent Similarity: 64.218      Percent Identity: 32.938

alignment block:
US-09-980-881-1 x US-09-171-945-125  ..

Align seg 1/1 to: US-09-171-945-125 from: 1 to: 716

24  CTTTCAGAGCCTTTCAGATCCTTTCAGCCATGTTCTCTTCCTGTCAGAGCA 73
   ::  ::::|  |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
1  MetLeuAlaLeuValLeuValIleThrValAlaLeuAlaSerAlaHisIle 17
74  TGTCTTCGGG...TTCCAGAGTGGCCAGTTCAGCTGCTTCCTTCAGAA 120
   |  ||:::  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
17  sglglglulhisrphgluglglululysvalIlePheArgValAlaAsnValGlu 34
121 CCTCTAGGCAAGTTCAGTTCACAGATCTTACTACACATATGAGATT 170
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
34  spGIuAsnHisIleAsnIleIleArgGluLeuAlaSerThrThrGlnIle 50
171 GTTCTCTGGCAGCCGGTAACAGCTGACCTTATGTGTGACAGAAAAACAAGT 220
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
51  AspPheThrIlysrProAspSerValIleThrGlnIleLysProHisSerThrVa 67
221 CCATTTTGTGTAATGCATCTGATGTCGACAAATGTAAGCCCATTTAA 270
   |  ||  ||:::  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
67  IAspPheArgValIlyslagIulnsrphrValIleThrValGluAsnValLeu 84
271 ATCTGAGCGGAATTCATGCAGTCTCTTGTCGACACGCGGAAGATCTT 320
   ::  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
84  yslslnsnIuLeuGlnIlyrLysValIleuIleSerAsnLeuArgAsnVal 100
321 ATTCAACACGACAGATTCACAAACGACAGTCAGCCCGGAGCTCCGCAATC 370
   ::::|  ||:::  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
101  ValGluAlaGlnPheAspSerArgVal.....ArgAlaThrGlnIle 114
371 GTACTATGACAGATTCATCACTCACTAAATGAAATCTATTCCTGGATGAAT 420
   |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
114  sSerIyrgIulysIyrAsnIlysrpIulThrIleGluAlaIleThrGlnG 131
421 TTATTACTGAGAGGATCCTGATATGCTTACAAAAATCCCATGATGCC 470
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
131  IuValAlaIleThrGluAsnProAlaIleuIleSerArgSerValIleGlyThr 147
471 TCATTTGAGAGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAACACA 520
   ::::|  |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
148  ThrPheGlnGlyArgAlaIleIyrlleuLeuLysVal...GlyLysAlaG 163
521 AACAGCCAAAAATGCCATATGATGATCTGTGGAATCCATGCCAGAGAT 570
   ||  |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
163  yglAsnIlysrProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 180
571 GGAATCTCTGCTTCTGCTTCTGCTGATCA..... 602
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
180  rIleSerProAlaPheCysGlnIlyrPheValArgGluAlaValArgThr 196
602  ..... 602

197  TyrgIyArgGluIleGlnValIleThrGluLeuLeuAspLysLeuAspPheTy 213
603  .....GCCCAT.....A 609
213  rValLeuProValIleuAsnIleAspGlyIyrlleTyrlThrIlyrThrLys 230
610  ATCGAATGTGGAGAAAGAACCGTCTTTCTATGCGAACAATCATGTGATC 659

```

```

:::|||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
230  eraIyPheThrPargLysThrArgSerThrHisThrGlySerSerCysIle 246
660  GGAACAGACCTGATAGACACTTGTCTCCAAACACACTGTGTGAGAGAG 709
   |||||  ||  |||||  ::  |||||  ||
247  GlyThrAspProAsnArgAsnPhe...AspAlaGlyIlyrPcysGluIleG 262
710  TGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGACATTATTCGTGAT 759
   |||||  ::  ||:::  |||||  |||||  |||||  |||||  |||||
262  yAlaSerArgAsnProCysAspIulThrIyrcysGlyProAlaAlaGlu 279
760  CAGAACCAAGAGTGAAGCGAGTGGCTAGTTCCTTGAGAAATAATATCAAC 809
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
279  ergIuLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295
810  CAGATTAACCATATACATCAGCATGATTCATTCCTCCAGCATATAGTGT 859
   ::::|  |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
296  SerIleLysAlaIyrlleuThrIleHisSerIySerGlnMetIleIy 312
860  TCCATATTCCTATACACAGAAATAAAGCAACATGAGAACTGTCTC 909
   |||||  ::  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
312  rProIySerIyAlaIyrlLysLeuGlyLysAsnAsnAlaGluLeuAsnA 329
910  TAGTAGCAGTGAAGCAGTTCGTCTATTCACAAACTATTAATAATAC 959
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
329  IeLeuAlaLysAlaThrValLysGluLeu...AlaSerLeuHisGlyThr 344
960  AGGTATACACATGCGCATGCGTCAGAAACCTTATACCTAGCTCCTGGAG 1009
   ::::|  |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
345  LysIyThrIyrlLysProGlyAlaIleThrIleIyIyProSerAlaGlyTh 361
1010  TGGGAGCAGATTTGATCTATGATTTGGGCATCAATATTCGTTTACATC.. 1057
   ::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
361  rSerLysAspIlyrAlaIyrcyspIingIlyIleArgIySerPheThrPheG 378
1057  ..... 1057

378  IuLeuArgAspThrGlyArgIyGlyPheLeuLeuProGluSerGlnIle 394
1058  AAACCCACCTGTAGAGAACCTTGTGCGCTGTCTTAAATATACCTTGCA 1107
   ::  |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
395  ArgAlaIleThrCysGlnGluThrPheLeuAlaIleLysIyValAlaSerIy 411
1108  TGTCAATTAGGAATGTT 1123
   ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
411  rValLeuGlnHisLeu 416

seq_name: /cgn2_6/plodata/1/1aa/5B_COMB.pep:US-08-782-760-6
seq_documentation_block:
; Sequence 6, Application US/08782760
; Patent No. 5948668
; GENERAL INFORMATION:
; APPLICANT: Hartman, Jacob
; APPLICANT: Fulga, Netta
; APPLICANT: Mendelovitch, Simona
; APPLICANT: Gorecki, Marian
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/782,760
FILING DATE: 13-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/378,233
FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0336/43847
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-782-760-6

```

```

Alignment Scores:
Quality: 607.50      Length: 305
Ratio: 3.148         Gaps: 6
Percent Similarity: 63.279      Percent Identity: 41.967

```

```

Alignment Block:
US-09-980-881-1 x US-08-782-760-6 ..

```

```

Align seg 1/1 to: US-08-782-760-6 from: 1 to: 307

```

```

360 GCGTCGCGATGCTACTATGACAGTATCATCTAAATGAATCATATTC 409
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1  AlaserGlyHisSerTyrThrLysTyrHisAsnTPrgLutThrIleGluAl 17
410 TTGATATGATTTATTAAGTATGAGAGCATCTGATATGCTTACAAAATTC 459
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 atPrlleGlnGlnValAlaThrAspAsnProAspLeuValThrGlnSerV 34
460 ACATGATGATCCATTTGAGAGATGACCCATGATGATGATGATGATTC 509
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 alIleGlyThrThrPheGlnGlyArgAsnMetTyrValLeuLysIle... 49
510 GGAAGAAGAACAGACAGCAAAATGATGATGATGATGATGATGATGAT 559
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 GlyLysThrArgProAsnLysProAlaIlePheIleAspCysGlyPheI 66
560 TGCCAGAGATGATCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 602
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 salArGlnLutPrlleSerProAlaPheCysGlnTrpPheValArgGlu 83
602 ..... 602
83 laValArgThrTyrAsnGlnGlnIleHisMetLysGlnLeuLeuAspGlu 99
603 .....GGCCAT..... 608
100 LeuAspPheTyrValLeuProValValAsnIleAspGlyTyrValTyr 116
609 .....AATCGAATGTGAGAAAGACCGTTCTTTCTATGCAACA 648
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 rTrPThrLysAspArgMetLtrPArgLysThrArgSerThrMetAlaGly 133
649 ATCATGTGATGGAGACAGACGCTGATATGCTGCTGCTGCTGCTGCTG 698
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 eSerCysLeuGlyValAspProAsnArgAsnPhe...AsnAlaGlyTyr 148
699 TGTGAGAGAGTGCATCCAGTTCCTCATCTCGAAACCTACTGTGACT 748
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 CysGlnValGlyAlaSerArgSerProCysSerGlnThrTyrCysGlyP 165
749 TTATCTGTGATGAGAACAGAGAGGAGGCTAGTTCTTTCAGAA 798

```

```

165 oAlaProGlnSerGlnLysGlnThrLysAlaLeuAlaAspPheIleArg 182
799 GAAATATCAACCAATTAACCATACATACACATGATCATCTATCCAG 848
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
182 snAsnLeuSerThrIleLysAlaTyrLeuThrIleHisSerTyrSerGln 198
849 CATATAGTGTTCATATCTATATACAGAGTAAAGCAAAACCATGA 898
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 MetMetLeuTyrProTyrSerTyrAspTyrLysLeuProGluAsnTyrG 215
899 GAACCTGTCTCTAGTACCAAGTGAAGCAGTTTCGTCATTTACAAACTA 948
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 uGlnLeuAsnAlaLeuValLysGlyAlaAlaLysGlnLeu...AlaThrL 231
949 GTAAATACAGGATATACATACATGCGCATGATGATGATGATGATGAT 998
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 euHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 247
999 GCTCCTGAGGTGGGAGCATTTGATCTATGATTTGGCATCAATATTC 1048
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 AlaAlaGlyLysSerAspAspTyrPserTyrAspGlnGlyIleLysTyr 264
1049 GTTATCATC..... 1057
264 rPheThrPheGlnLeuArgAspThrGlyPhePheGlyPheLeuLeuPro 281
1058 .....AAACCACTGTAGAGAGCTTTGCCGCTGTCTATAA 1096
281 lUserGlnLeuArgGlnThrCysGlnGlnLutThrMetLeuAlaValLys 297
1097 ATAGCTTGCGATGTC 1111
298 lLeAlaAsnTyrVal 302
seq_name: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:PCT-US96-00995-6
seq_documentation_block:
: Sequence 6, Application PC/US9600995
: GENERAL INFORMATION:
: APPLICANT: Bio-Technology General Corp.
: TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/00995
: FILING DATE: 25-JAN-1996
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 0336/43847-A-PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 308 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein

```


This Page Blank (uspto)

[illegible]

```

seq_name: p1r2:A32129
seq_documentation_block:
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C.Species: Rattus norvegicus (Norway rat)
C.Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
CAccession: A32129; S17543
R.Clauser, E.; Gaidell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 265, 17837-17845, 1988
A.Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comp
A.Reference number: A92693; MUID:89034324
A.Accession: A32129
A.Molecule type: DNA
A.Residues: 1-415 <CLAB>
A.Cross-references: GB:M23959; GB:J04041; NID:g203293; PIDN:AAA40872.1; PID:g203295
R.Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A>Title: Distribution of manganese in rat pancreas and identification of its primary
A.Reference number: S17543; MUID:91378950
A.Accession: S17543
A.Molecule type: Protein
A.Residues: 109-116, 'X', 118-130 <KOD>
C.Genetics:
A.Gene: CPB
A.Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C.Keywords: hydrolase; metallo-carboxypeptidase
F.354,376/Active site: Tyr, Glu #status predicted

alignment_scores:
Quality: 695.50 Length: 419
Ratio: 2.644 Gaps: 8
Percent Similarity: 62.766 Percent Identity: 37.709

Alignment_block:
US-09-980-881-1 x A32129 ..

Align seg 1/1 to: A32129 from: 1 to: 415

18 ATGAAGCTTTGCAGCGCTTGCAGTCCTTGACCATTGTCTCTGTA 67
||| ||| ||| ||| |||
1 MetLeuLeuLeuAlaValSerValAlalaValAlaHisAsnLac 17
GGACATGCTCCGCCCTCCAGTAGGCCAAGTTACTGCTGCTCTCTA 117
||||| |
68 GCAGCATGCTCCGCCCTCCAGTAGGCCAAGTTACTGCTGCTCTCTA 117
||||| |
17 ugnlhis.....PheaspGlyAsnArgValIyArgValISerValH 31
||||| |
118 GAACCTTAGGCACAAGTTCAAGTTCTACAGAATCTTACTACAACAAATATG 167
:: :: ::::::::::::||| |||
31 ISGLYGIAspHisValAsnLeuIleGlnGluLeuAlaAsnthrlgslu 47
||||| |
168 ATTGTTCTCGTGGACCGGTACAGCTGACCTTATGTCGAAGAAAAACA 217
||| ||| ||| ||| |||
48 IleaSrpethrlPylsrProaspserAlathrelvalLysProleuthrh 64
||||| |
218 AGTSCATTTTTTGTAAATGCATGTGATGTGCACAATGTGAAGCCCAT 267
||| ||| ||| ||| ||| ||| ||| |||
64 rValaSrphenHisValLysAlaGluaSpValAlaAspValaluaSnPhel 81
268 TAAATGTAGCSGGAATTCATGACAGCTGTCTGCTGGCAGACSTGGAAAT 317
||||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
81 euGluGluAsnGluValHisTyrgluValLeuIleSerAsnValaArgsn 97
318 CTATTATCAACGACAGATTCCAACGACAGTACAGCCCCGAGGCTCGC 367
::::::::::::::::::||| |||
98 AlaLeuGluSerGlnPheaspserHisThr.....ArgAlaSerGel 111
368 ATCGTACTATGACAGTATCACCTGAATAATGAATGTAATCTATTCGATAG 417
||| ||| ::|||::: :: :: ||| ::|||:::
111 yHisSerVthrLyssrYasnLvstPgIuThlllgJuValatrrdileg 128
```



```

166 gLysAlaIlePheMetAspCysGlyIleHisAlaArgLutPrIleSerP 183
580 CTGCTTCTGCTTGTGCTCATTA.....GGCCAT 608
|||||.....|
183 rolaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199
609 AAT.....|
200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPhetYrValLeuPr 216
612 .....CGAATG 618
216 oValPheAsnValAspGlyTyrIleTrpSerTrpTrpGlnAspArgMet 233
619 GGAGAAAGAACCGTCTTCTATCGAACAAATCGATCGAGAACAGAC 668
|||||.....|
233 rPrArgLysAsnArgSerArgAsnGlnAsnSerThrCysIleGlyThrAsp 249
669 CTGAATAGCAACTTGTCTCCAAACACTGCTGAGAGAGCGTCATCCAG 718
|||||.....|
250 LeuAsnArgAsnPhe...AspValSerTrpAspSerSerProAsnThrAs 265
719 TTCCTCATGCTCGAAGACTGCTGAGACTTATCCGAGTCAGAACCCAG 768
::: ||| ::::| ||| |||||
265 nysProCysLeuAsnValTyrArgGlyProAlaProGlnSerGluLysG 282
769 AAGTGAAGCAGATGCTAGTCTTGAAGAAATATCAACAGATTAA 818
||:|||||:|||||:|
282 LurHisAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
819 GCATATACAGCATGATCATCTATCCCAATAGCTGTTCCTCATATTC 868
|||||:|||||:|||||:|
299 AlarTyrIleThrPheHisSerTyrSerGlnMetLeuIleProTyrG 315
869 CTAATACAGAGTAAAGCAAGAACCATGAGAACTGCTCTAGTAGCCA 918
:|||||:|
315 yTyrThrPheLysLeuProProAsnHisGlnAspLeuLeuValAlaAla 332
919 GTGAGACCATGCTGCTATTTGACAAATAGTAAATACAGATATCA 968
|||||:|||||:|
332 rGleIleThrAspAlaLeu...SerThrArgTyrGluThrArgTyrIle 347
969 CATGCCATGCGCTCAGAAACCTTATCTAGCTCTGAGGTGGAGACGA 1018
::: ||| ::::| ||| |||||
348 TyrGlyProIleAlaSerThrIleTyrLysThrSerGlySerLeuAs 364
1019 TTGGATCATGATTTGGCATCAATATTCGTTAC.....ATCAAAACCACC 1066
|||||:|||||:|
364 pTyrValTyrAspLeuGlyIleLysHisThrPheAlaPheGluLeuArg 381
1055 .....ATCAAAACCACC 1066
381 sPlyGlyLysSerGlyPheLeuLeuProGlnSerArgIleLysProThr 397
1067 TGTAGAGAGCTTTTGCCGCTGCTCTATAAATAGCTTGGCATGCTTAG 1116
|||||:|||||:|
398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeu 414
1117 GAAT 1120
: |||
414 sAsn 415

seq_name: p1r2:A43929
seq_documentation_block:
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N:Alternate names: mast cell carboxypeptidase A3
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence, revision 31-Dec-1993 #text_change 22-Jun-1999
C:Accession: A43929, A39246, A45759
R:Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A:Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A

```

```

A:Reference number: A43929; MUID:92105393
A:Accession: A43929
A:Molecule type: DNA
A:Residues: 1-417 <REV>
A:Cross-references: GB:W3716
A:Experimental source: mast cell
A:Note: The authors translated the codon GCG for residue 211 as Thr
R:Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Sugarbaker, D.J.; Austen, K.F.; Serafi
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A:Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and compari
A:Reference number: A39246; MUID:90083291
A:Accession: A39246
A:Molecule type: mRNA
A:Residues: 1-417 <REV>
A:Cross-references: GB:M2717; NID:9179933; PID:AAA35652.1; PID:9179934
R:Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.O.
J. Clin. Invest. 83, 1630-1636, 1989
A:Title: Human mast cell carboxypeptidase. Purification and characterization.
A:Reference number: A45759; MUID:89214692
A:Accession: A45759
A:Molecule type: protein
A:Residues: 110-137 <GOL>
C:Genetics:
A:Gene: GDB:CPA3
A:Cross-references: GDB:125211; OMIM:114851
A:Map position: 3q21.3-3q25
A:Introns: 23/2; 48/3; 90/2; 124/3; 158/3; 192/3; 229/3; 260/1; 327/3; 356/1
C:Keywords: carboxypeptidase
C:Superfamily: carboxypeptidase
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-109/Domain: activation peptide #status predicted <ACT>
F:110-417/Product: carboxypeptidase A, mast cell #status predicted <MAN>
F:176-179,304/Binding site: zinc (His, Glu, His) #status predicted
F:245-268/Disulfide bonds: #status predicted
F:356,378/Active site: Tyr, Glu #status predicted

alignment_scores:
Quality: 647.50 Length: 427
Ratio: 2.398 Gaps: 10
Percent Similarity: 63.232 Percent Identity: 33.488

alignment_block:
US-09-980-881-1 x A43929 ..

Align seg 1/1 to: A43929 from: 1 to: 417

6 ATTGCGTGGGATGAAGCTTTGCAAGCTTGCAGCTTCCTGTACCAATTGT 55
::: |||||::: ::|||:::
5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle...AlaProVal.. 19
20 .....ArpPheAspArgGlyLysValPheA 28
56 TCTCTTCTGTGAGCAGCATGTCTGCGCTTCCAGAGATGGCCAAATCTAG 105
|||||:|||||:|
20 .....ArpPheAspArgGlyLysValPheA 28
106 CTGCTCTTCTGTAAGCCTTGCAGCAAGTTCAAGTTCTACAGAAATCTACT 155
::: |||||:|||||:|
28 rGValLysProGlnAspGlyLysGlnAlaAspIleIleLysAspLeuAla 44
156 ACAACATATGATGATTTGCTCTGCGCAGCGGTAACAGCTGACCTATTTGT 205
||| |||||:|
45 LysThrAsnGluLeuAspPheThrPyrTrpGlyAlaThrHisHisValAla 61
206 GAAGAAAAAAGCAAGTCCATTTTGTAAATGATCATCTGATGCGACAATG 255
: :: ||||| ||| |||||:|
61 AlaIleAsnMetMetValAspPheArgValSerGluLysGlnAlaIle 78
256 TGAAGACCCATTTAAATGATGAGCGGAATTCATGCAAGTGTCTGCTGGCA 305
|||||:|||||:|
78 LeuGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuHis 94
306 GACGTGGAAGATCTTATTCACAGCAGATTCC...AACGACACAGTCAG 352
|||||:|||||:|

```


alignment_block:
US-09-980-881-1 x CPBOB ..

Align seg 1/1 to: CPBOB from: 1 to: 306

```

375 TATGACAGATATCAGTCACATAAATGAAATCTATCTTGGATAGCAATTTAT 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6 TyrglulysTyraasnAsnTrpGluThrIleGluAlaTrpThrGluGlnVal 22
425 AACTGAGAGCATCTGATGATGCTTACAAAATTCACATGATGATCTCAT 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 AlaSerGluAsnProAsnProAsnLeuSerAlaIleGlyThrThrP 39
475 TTGAGAAGTACCCACTCTATGTTTAAAGTTTCTGGAAAAAGAACAAACA 524
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 heLeuGlyAsnThrIleTyrlLeuLeuVal...GlyLysProGlySer 54
525 GCCAAAAATCCCATATGATGATGATGATGATGATGATGATGATGATGAT 574
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 AsnLysProAlaValPheMetAspCysGlyPheHisAlaArgLysIrrPIL 71
575 CTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 eSerProAlaPheCysGlnTrpPheValaArgLysAlaValaArgThrTyrg 88
602 ..... 602
88 LysArgLysIleHisMetThrGluPheLeuAspLysLeuAspPheTyrlVal 104
603 .....GCCCAT.....AATCG 613
105 LeuProValaValAsnIleAspGlyTyrlLeuTyrlThrTrpThrAsnArg 121
614 AATGTGAGAGAAGAACCGCTCTTCTGATGCGAACATCATGATGATGCGAA 663
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 gMetTrpArgLysThrArgSerThrArgAlaGlySerSerCysThrIrr 138
664 CAGACCTGAATAGCAACTTGTCTCCAAACACTGCTGCGAAGAGTGA 713
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 hAspLeuAsnArgAsnPhe...AspAlaGlyTrpCysSerIleGlyAla 153
714 TCCAGTTCCCATGCTCGGAAACCTACTGAGACTTTATCTGATGATGACA 763
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 SerAsnAsnProCysSerGluTrpTrpCysGlySerAlaIleGlySerG 170
764 ACCGAGAGTGAAGGACAGTGGCTAGTTCTTGAGAGAAATATCAACAGA 813
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 uLysGluSerLysAlaValaIleAsnProIleArgAsnHisLeuSerSer 187
814 TTAAAGCATATCATCAGCATGATCATATCTCCAGCATATAGTGTTCACA 863
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 LeuSerAlaTyrlLeuThrIleHisSerLysSerGlnMetLeuTyrlPro 203
864 TATTTCTATATCAGCAAGTAAAGCAAGCAAGCAAGCAAGCAAGTGTCTAGT 913
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 TyrlSerTyrlAspTrpTyrlLysLeuProLysAsnAsnValGluLeuAsnThrLe 220
914 AGCCAGATGAGCAGTTCGCTATTTGACAACAACTAGTAAATAATCAGAGT 963
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 uAlaLysGlyAlaValaLysLysLeu...AlaSerLeuHisGlyThrThrT 236
964 ATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGAGAGTGG 1013
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 ySerTyrlGlyProGlyAlaIleThrIleTyrlProAlaSerGlyGlySer 252
1014 GAGGATGGATCTATGATTTGGGCAACAATATCTCTTACATC..... 1057
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 AspAspTrpAlaTyrlAspGlnGlyIleLysTyrlSerPheThrPheGluLe 269
1058 .....AAC 1061
269 uArgAspLysGlyArgTyrlGlyPheValaLeuProGluSerGlnIleGlnP 286

```

```

1062 CCACCTGTAGAGAACCTTTGGCCCGCTGTCTCTAAATAATGCTTGCATGTC 1111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 rofhrCysGluGluThrMetLeuAlaIleLysTyrlValThrSerTyrlVal 302
1112 ATTAGCAATGTT 1123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 LeuGluHisLeu 306
seq_name: p1r2:A38395

```

```

seq_documentation_block:
mast cell carboxypeptidase (EC 3.4.-.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C:accession: A38395; A33118
R:cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
  Biochemistry 30, 648-655, 1991
A:title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme a
A:reference number: A38395; MOID:91105153
A:accession: A38395
A:status: preliminary
A:molecule type: protein
A:residues: 1-309 <COL>
R:cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
  submitted to the Protein Sequence Database, August 1990
A:reference number: A33118
A:accession: A33118
A:status: preliminary
A:molecule type: protein
A:residues: 2-309 <COL>
C:superfamily: carboxypeptidase
C:keywords: hydrolase
F:248,270/Active site: Tyr, Glu #status predicted

```

alignment_scores:
Quality: 567.50 Length: 303
Ratio: 2.925 Gaps: 6
Percent Similarity: 64.026 Percent Identity: 39.604

alignment_block:
US-09-980-881-1 x A38395 ..

Align seg 1/1 to: A38395 from: 1 to: 309

```

375 TATGACAGATATCAGTCACATAAATGAAATCTATCTTGGATAGCAATTTAT 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8 TyrlAlaLysTyraasnAsnTrpAsnLysIleValaSerTrpThrGluLysWe 24
425 AACTGAGAGCATCTGATGATGCTTACAAAATTCACATGATGATGATGATGAT 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24 tValGluLysHisProGluMetValaSerArgIleLysIleGlySerThrArg 41
475 TTGAGAAGTACCCACTCTATGTTTAAAGTTTCTGAAAAAGAACAAACA 524
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 alGluAspAsnProLeuTyrlValLeuLysIle...GlyArgLysAspLys 56
525 GCCAAAAATCCCATATGATGATGATGATGATGATGATGATGATGATGAT 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 GluTrpLysAlaIlePheMetAspCysGlyIleHisAlaArgLysIrrPva 73
575 CTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 lSerProAlaPheCysGlnTrpPheValaArgLysAlaLysSerTyrg 90
604 GCCATTAAT..... 611
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 LysAsnAsnIleMetThrLysLeuLeuAspArgMetAsnPheTyrlVal 106
612 .....CG 613
107 LeuProValaPheAsnValaAspGlyTyrlIleTrpSerTrpThrLysAspArg 123

```

```

614 AATGTGAGAAAGAACCGTTCTTTATGCGAAACATCATGCAATGGGAA 653
|||||
123 gmetrparglysasnarlyserlysasnproserSerThrCys116GlyT 140
|||||
664 CAGACCTGAATAGCACTTTGCTCCAAACACGTCGTGCGAAGGCGCA 713
|||||
140 hrAspleuasnarganphe...AspValSerThrAspSerSerProasn 155
|||||
714 TCCAGTTCTCATGCTCGGCAACCTGAGCTTATTCCTGAGTCAGA 763
|||||
156 ThrAspnProCysLeuSerValTyrArgGlyProAlaProGluSerG 172
|||||
764 ACCAGAGGAAAGGACGTCGCTAGTTCTTGAGAAATAATACACAGA 813
|||||
172 uLysGluThrLysAlaValThrAsnPhelIeArgSerHisLeuAsnSer 189
|||||
814 TTTAAACATACATCAGCATGATTCATCTCCAGCATATAGTTTCCCA 863
|||||
189 lelysalatyrIleThrPhelIeSerTyrSerGlnMetLeuLeuPhePro 205
|||||
864 TATTCCTATACACGAACTAAAGCAACCATGAGAACCTGCTCTAGT 913
|||||
206 TyrGlyTyrThrIleLysLeuProProAsnHisGlnAspLeuLysVa 222
|||||
914 AGCCAGTGAAGCAGTCGTCCTATTGACAAACTACTAAAAATACACAG 963
|||||
222 lAlaArgIleAlaThrAspValLeu...SerSerArgTyrGluThrArg 238
|||||
964 ATACATGATGCCATGCGCTCAGAAACCTTATACCTGCTGAGAGTG 1013
|||||
238 yrlIeTyrGlyProIleAlaSerThrIleTyrLysThrSerGlySerSer 254
|||||
1014 GACGATGGATCTGATGTTTGGGATCAAAATATTCGTTTAC..... 1054
|||||
255 LeuAspTrpAlaTyrAspLeuGlyIleLysHisThrPheAlaPheGlu 271
|||||
1055 .....ATCAAC 1061
|||||
271 uArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLasp 288
|||||
1062 CCACCTGTGAGAGAGCTTTGGCGCTCTCTAAATACCTTGGCAATGC 1111
|||||
288 rOthCysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIle 304
|||||
1112 ATTAGGAAT 1120
|||||
305 LeuLysHis 307
|||||
seq_name: pLr2:A32128
seq_documentation_block:
cardoxypeptidase A2 (EC 3.4.17.15) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C:Accession: A32128
R:Gardell, S.J.; Craik, C.S.; Clausen, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rut
J. Biol. Chem. 263, 17828-17836, 1988
A:Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and ex
A:Reference number: A32128; M01D:89034323
A:Accession: A32128
A:Molecule type: mRNA
A:Residues: 1-417 <GAR>
A:Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956.1; PID:g
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:356,378/Active site: Tyr, Glu #status predicted
alignment_scores:
Quality: 553.00 Length: 409
Ratio: 2.194 Gaps: 13
Percent Similarity: 61.614 Percent Identity: 33.496

```

```

alignment_block:
US-09-980-881-1 x A32128 ..
Align seg 1/1 to: A32128 from: 1 to: 417
18 ATGACGTTTGCACCTTCGCTGCTTACCCATTTGTTCTTCTGTGA 67
|||||
1 MetArgLeuThrLeuLeuAlaIleLeuGlyTyrIleTyrCysG 17
68 GCAGCATGCTTCGCGTCCAGAGTGGCCAGTTCTTACGCTCTCCCA 117
|||||
17 nGlu.....ThrPheValGlyAspGlnValLeuGluIleProS 31
118 GAACCTTGAGCAAGTTCAATCTACAGAACTTACTACACA..... 161
|||||
31 erHisGluGluGlnIleArgThrLeuLeuGlnLeuGlnAlaGluHis 47
162 TATGACATTTGTTCTTCGCGCGGTACAGCTGACCTTATTTGTGA 211
|||||
48 LeuGluLeuAspPheTyrLysSerProThr.....IleProG 60
212 AAACAGTCCTATTTTGTAAATGATCATGATGTCGCAATGTGAAG 261
|||||
60 yGluThrValHisValAlaGlyValProPheAlaSerIleGlnAlaValLysV 77
262 CCCATTTAAATGTGACCGGAAATTCATGCAAGTCTTGTGCGACGAG 311
|||||
77 alPheLeuGluSerGlnGlyIleAspTyrSerIleMetIleGluAspVal 93
312 GAAGATCTTATTCACAGCAGATTTCCAC.....GACACACTGAG 352
|||||
94 GlnValLeuLeuAspGlnGluArgGluMetLeuPheAsnGlnGlnAr 110
353 CCCCAGGCGCTCGCATCTACTATGACAGATGATCACTCACTAAATGA 402
|||||
110 gGluArgGlyGlyAsnPheAsnPhelValAlaTyrHisThrLeuGluLur 127
127 letyrgInGluMetAspAsnLeuValAlaGluAsnProGlyLeuValSer 143
403 TCTATTCCTTGATAGATTTATTAAGTACAGAGCATCTGATATGCTTAC 452
|||||
127 letyrgInGluMetAspAsnLeuValAlaGluAsnProGlyLeuValSer 143
453 AAATTCACATGATGATCCATTTGAGAAATGACCCACTATGTTTAA 502
|||||
144 LysValAsnLeuGlySerPheGluAsnArgProMetAsnValLeuLys 160
503 GGTTCCT...GGAAGAAGACAAACAGCCAAATATGATGATGATGACT 549
|||||
160 spHeSerThrGlyGlyAsp.....LysProAlaIleThrPheAspA 174
550 GTGGAATTCATGCGAGAGATGATCTCTGCTTCTGCTGTGG... 596
|||||
174 lagIylIleHisAlaArgGluTrpValThrGlnAlaThrAlaLeuTrpThr 190
596 ..... 596
191 AlaAsnLysIleAlaSerAspTyrGlyThrAspProAlaIleThrSerIle 207
597 .....TTCATA.....GGCC 606
207 uLeuAsnThrLeuAspIlePheLeuLeuProValThrAsnProAspGlyT 224
607 AT.....AATGAAATGGAAGAAACCGTTCTTTC 638
|||||
224 yValAlaPheSerGlnThrThrAsnArgMetTrpArgLysThrArgSerLys 240
639 TATGGAACAATCATTCATCGGAACAGCAGTGAATAGCACTTGTCTC 688
|||||
241 ArgSerGlySerGlyCysValGlyValAspProAsnArgAsnTrp...As 256
689 CAACACGTGCTGTGAGAGGTGATCCGATCCGATCATGCTGGAACCT 738
|||||
256 palAsnPhelGlyGlyProGlyAlaSerSerSerProCysSerAspSerT 273

```



```

256 TGAAGCCATTTAAATGTCAGCGAATTCATGCACTGTCTGTCGCA 305
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    75 allysvalpheleugluahisgllyeargyrarglilemetileglu 91
306 GAGCTGAGAGATCTTAT.....CAACAGAGATTCACAGCAC 346
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    92 Aspyvalginserleuleuaspulguglucgluimethphelaasergl 108
347 AGTCAGCCCCAGAGCTCCGATCGTAC...TATGAACATGATCACTAC 393
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    108 nserarglalaargserthrashanthrphenanthyralaethrthhshrl 125
394 TAAATGAATCTATCTGTCGATGAATTTTAACTGAGAGCATCCGAT 443
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    125 euaspuglulietyrasphmethasphleuvalalagluhshsprogln 141
444 ATGCTTACAAAATCCACATGATGATCTCATTTGAGAAATGACCATCTA 493
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    142 leuvalserlyseuglnilegllyargserlyrgluclyargproliety 158
494 TGTTTTAAAGTTTCTGGAAAAGAACACAGCCAAAATGCAATATGA 543
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    158 ryalteulysrpheser...Thrglyyserashanargproalailertprl 174
544 TTGACTGTGGAATCCAGAGCAAGATGATCTCTCTCTCTCTCTCTG 593
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    174 leaspugllyliehshseratrglultrillerghlnalatrghlyval 190
594 TGGTTCATA..... 602
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    191 TrpPhelalalyserphethrgluaspyrglyglinasproserpheth 207
602 ..... 602
207 ralaileuaspsermetasphlephleugluilevalthrashpro 224
603 .....GGCCATATCGAATGTGGAGAAAGAACCGT 632
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    224 spgllyphehalarphethrhisaserghashnrgyleutrpshlysthr 240
633 TCTTTATGCGAACAATATTCATGCGAAGACCTGATAGCAACT 682
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    241 servaltthrseraserleucysvalgllyalaspalashasnrgshnr 257
683 TGTCTCCAAACACTGCTGTGAGAGAGTCATCTCTCTCTCTCTG 732
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    257 P...AspAlaglyPhegllysalaglyalaseraserProCysserg 273
733 AAACCTACTGTGACTTTATCTGAGTCAAGACCAAGTGAAGGAGT 782
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    273 luthrlyrhisgllystlyrhalashsergluvalgluvallyssertile 289
783 GCTAGTTCTTGAGAAATATCAACAGATTAAAGATATACATGAGCAT 832
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    290 ValAspPhevallylsasphshsclysasn...PheylsAlaPheleuseril 305
833 GCATTCATCTCCAGCATATAGTGTTCATATTCCTATACAGAACTA 882
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    305 ehlsaserlyserghlnleuleuleuyltrporyrglytrthrhglshs 322
883 AAAGCAAAAGCATGAGAACTGTCTAGTACGAGCAAGCAAGTTCGT 932
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    322 erlileprohsprlysthrghlualeuashgluvalalalysserlavalglu 338
933 GCTATGACAAAATGTAATAATACAGAGTATACATGAGCATGAGCTC 982
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    339 Alaleu...lyserleuyltrghlytrserlyrlystlyserlilel 354
983 AGAAACCTTATACCTAGCTCCGAGAGTGGAGCATGATGATCTATGAT 1032
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    354 eThrthrleryrglnalaserghlyserlileasprlserlyrlyasng 371
1033 TGGCATCAAAATATTCGTTTACATCAAA..... 1060

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
371 lnglyliletyrserphethrphleugluuargrasprthrglyargtyr 387
1061 .....CCGACCTGTGAGAACGCTT 1080
388 GlyPheleuleuProalaserghlnleleprothrphalaglulthrtr 404
1081 TGGCGCTGTCTTAAATAGCTTGGCATGTCATGATGAGATGTT 1123
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    404 pleuglyvalleuthrlethlemetghlnshthrlashanashleu 418
seq_name: plr1:CPRTA

```

```

seq_documentation_block:
carboxypeptidase A (EC 3.4.17.1) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jun-1983 #sequence #revision 13-Jun-1983 #text-change 24-Sep-1999
C:Accession: A00911; B32129
R:Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Stranding, D.N.; Pictet,
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
A:Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization
A:Reference number: A00911; MUID:82105986
A:Accession: A00911
A:Molecule type: mRNA
A:Residues: 1-419 <QD>
A:Cross-references: GB:J00713; NID:G203364; PID:AAA40893.1; PID:G203365
R:Clausner, E.; Gardell, S.J.; Craik, C.S.; Macdonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A:Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comp
A:Reference number: A92693; MUID:85034324
A:Accession: B32129
A:Molecule type: DNA
A:Residues: 1-260, 'F', 262, 'W', 264-346, 'K', 348-419 <CLA>
A:Genetics:
A:Gene: CP1A
A:Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A>Note: the authors translated the codon GTC for residue 196 as Ile
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; z
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-110/Domain: activation peptide #status predicted <PRO>
F:111-419/Product: carboxypeptidase A #status predicted <MAT>
F:179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F:248-271/Disulfide bonds: #status predicted
F:358,360/Active site: Tyr, Glu #status predicted

```

```

alignment_scores:
Quality: 529.00 Length: 436
Ratio: 2.124 Gaps: 13
Percent Similarity: 57.110 Percent Identity: 32.339.

```

```

alignment_block:
US-09-980-881-1 x CPRTA ..

```

```

Align seg 1/1 to: CPRTA from: 1 to: 419

```

```

21 AAGCTTTGAGAGCTTGCAGTCTGTATGACCATGTTCTCTGTGAGCA 70
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    3 Argyleuileuleuserleuleuileuvalalalyssgllyasnglshs 19
71 GCATGCTTGGCGTTCCAGAGTGGCCAAAGTTGATGCTTCTCTAGAA 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    19 n.....PhevalGlyhisgluvalleuarglileseralalalaa 32
121 CCGTAGAGAGTTCAGTTCATGAGATCTTACATGATGATGATGAT 170
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    32 spglualaglinalglulnlysalgllysglualeuaspheuglnshs 48
171 GTTCTC.....TGGCAGCGGTAACAGCTGATGATGATGATGATGAA 214
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
    49 GluleuAspPheThrpargrargalalalarglialgllyleproliaspva 65

```

```

215 ACAAGTCATTTTGTGTAATGCAATGATGTCACAAATGTAAGCC 264
    :||| |||
    :||| |||
65 lAtgValProPhe.....ProSerIleGlnSerValIysAla 78
265 ATTAAATGAGCGGAATTCATGATGATGCTGTGGAGAGCGTGA 314
    :||| |||
    :||| |||
78 heLeuGluTyrHisGlyIleSerTyrGluIleMetIleGluAspValGln 94
315 GATCTTATT.....CAACAGCAGATTTCACAGCAGACAGCAGCC 355
    :||| |||
    :||| |||
95 leuLeuLeuAspGlnGluGlnGlnMetSerIlePheGlnAlaAspAl 111
356 CCGAGCCTCCGCAATGTCAC...TATGAAACGATTAATGCTACTAAATGAA 402
    :||| |||
    :||| |||
111 aLeuSerThrAspSerPheAsnTyrIleThrTyrHisThrLeuAspGlu 128
403 TCTATTCTGGATGAAATTATTAAGTGAAGGAGCTGCTGATATGCTTACA 452
    :||| |||
    :||| |||
128 lEtyGlnPheMetAspLeuLeuValAlaGlnHisProGlnLeuValSer 144
453 AAATTCACATTTGATCCTCATTTGAGAAAGTACCACCTGATGTTTAAA 502
    :||| |||
    :||| |||
145 lYstIleGlnIleGlyAsnThrPheGlnGlyIleProIleHisValIle 161
503 GCTTTCGGAAGAAACAAACAGCCAAATGCAATGATGATGATGCTGTC 552
    :||| |||
    :||| |||
161 sPheSer...ThrGlyGlyThrAsnArgProAlaIleTyrIleAspThrG 177
553 GAATTCATGCGAGAGAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTA 602
    :||| |||
    :||| |||
177 lYleHisSerArgIleTyrValThrGlnAlaSerGlyValTyrPheAla 193
603 GGCAT..... 608
194 lYstValThrIysAspTyrGlyGlnAspProThrPheThrAlaValle 210
608 ..... 608
210 uAspAsnMetAspIlePheLeuGluIleValThrAsnProAspGlyPhea 227
609 .....AATGCAATGAGAGAAAGACCGTCTTCTCAT 641
    :||| |||
    :||| |||
227 lAtyThrHisLysThrAsnArgMetTyrArgLysThrHisSerHisThr 243
642 GCGAACAATGATGTCATGCGAACAAGACCTGAATAGCAATTGCTGCCA 691
    :||| |||
    :||| |||
244 GlnIleSerLeuCysValGlyValAspProAsnArgAsn..... 256
692 ACACTGG.....TGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
    :||| |||
    :||| |||
257 ....TTPAspAlaGlyLeuGlyLysAlaGlyIleSerSerAsnProCys 272
730 CGAACAACCTACTGTGACTTATCTGAGTGAAGAACCGAGTGAAGCA 779
    :||| |||
    :||| |||
272 eArgIleTyrArgGlyLysPheProAsnSerGlyValGlyValLysSer 288
780 GTGGCTAGTTTCTTGAGAAAGAAATATCAACCGATTAAGCATACATG 829
    :||| |||
    :||| |||
289 lIleValAspPheValThrSerHisGlyAsn...lIleLysAlaPheIleSe 304
830 CATGCAATTCATACCCCAAGCATATAGTGTTCATATTCCTATACAGCA 879
    :||| |||
    :||| |||
304 rIleHisSerTyrSerGlnLeuLeuTyrProTyrGlyTyrHisGrg 321
880 GTAAGAACAAGACCATGAGCAAGTGTCTGTAGTACAGAGCAAGCACT 929
    :||| |||
    :||| |||
321 lUrProAlaProAspGlnAlaGlyLeuAspGlnLeuAlaIleLysSerAlaVal 337
930 CGTGCCTATGCAAAACTAGTAAATATACAGATACATACATGCGCATG 979
    :||| |||
    :||| |||
338 ThrAlaLeu...ThrSerLeuHisGlyThrGluPheLysTyrGlySerI 353
980 CTCAGAAACCTTATACCTAGCTCTGGAAGGTGGGAGCATTTGATCTATG 1029

```

```

353 eIleAspThrIleTyrGlnAlaSerGlySerThrIleAspTyrThrTyr 370
1030 ATTGGCATCAAAATATTCGTTTACATCAAAAC..... 1061
    :||| |||
    :||| |||
370 eArgGlnGlyIleLysTyrSerPheThrPheGlnLeuAlaAspThrGlyLeu 386
1062 .....CCACCTGTAGAGAACGCTTGTGCCCTGTCTCTAAATA 1099
    :||| |||
    :||| |||
387 ArgGlyPheLeuLeuProAlaSerGlnIleIleProThrAlaGlnGlu.. 402
1100 GCTTGGCATGTCATTTAGCAATGTTTAAATGCCCTGATTTTATCATTTGCG 1149
    :||| |||
    :||| |||
403 .....ThrTyrPheAlaLeuLeuThrIleMetAspHisThrValLysH 417
1150 TTCGGTAT 1157
    :||| |||
417 lSProTyr 419
seq_name: p1r2:A56171
seq_documentation_block:
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N/Alternate names: pancreatic carboxypeptidase A2
C/Species: Homo sapiens (man)
C/Date: 28-Apr-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999
C/Accession: A56171; S02809; S71395
R/Catalsus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puljserver, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A/Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. CD
A/Reference number: A56171; M0ID:95204457
A/Accession: A56171
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-417 <CAT>
A/Cross-references: GB:019977; NID:9790226; PID:AAA74425.1; PID:9790227
A/Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue
R/Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A/Title: Purification and properties of five different forms of human procarboxypepti
A/Reference number: S02809; M0ID:89153096
A/Accession: S02809
A/Molecule type: protein
A/Residues: 'S',18-36,'N',38-43 <PAS>
R/Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Wal
Arch. Biochem. Biophys. 332, 8-18, 1996
A/Title: Expression and characterization of human pancreatic preprocarboxypeptidase A
A/Reference number: S71394; M0ID:96400327
A/Accession: S71395
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-301,'T',303-338,'SR',341-417 <LAD>
C/Genetics:
A/Genes: GDB:CPA2
A/Cross-references: GDB:125230; OMIM:600688
A/Map position: 7q32-7qter
C/Superfamily: carboxypeptidase
C/Keywords: hydrolase; metallo-carboxypeptidase; zinc
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-110/Domain: activation peptide #status predicted <ACP>
F:111-417/Product: carboxypeptidase A2 #status predicted <MAP>
F:177-180,304/Binding site: zinc (His, Glu, His) #status predicted
F:246-269/Disulfide bonds: #status predicted
F:356,378/Active site: Tyr, Glu #status predicted
alignment_scores:
Quality: 508.00 Length: 394
Ratio: 2.153 Gaps: 11
Percent Similarity: 59.898 Percent Identity: 32.487
alignment_block:
US-09-980-881-1 x A56171 ..

```

Align seg 1/1 to: A56171 from: 1 to: 417

```

72 CATGCTTC.....GCCTTCAGAGTGGCCAAAGTTCTAGTGGCTCT 112
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13 HisIeTyrcysLeuGluThrPheValIglYAspIlnValLeuGluIleVa 29
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 TCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAACTTACTACAAACAT 162
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 IProSerAsnIlnGluIlnIleYAsnIlnLeuGlnIlnLeuIlnIlnG 46
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 ATGAGATTGTCTC.....TGGCAGCGGTAAGAGCTGACGACTATGTG 206
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46 IlnHisIleGlnLeuAspPheTrpIlySerProItr.....Thr 58
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 AAGAAACAAACATCATTTTCTTAATGATCATGTGATGCGAAATGT 256
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 ProIgluThrIlnHisValArgValProPheValAsnValIlnAlaVa 75
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
257 GAAAGCCCATTTAAATGTGAGCGGAATTCATGCAATGTCTGTGTCGACG 306
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 IlyValPheLeuGlnIlnSerGlnIlyIleAlaIlySerIleMetIleGlu 92
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 ACGGGAAGATCTTATCAACAGCAGATTCCAAAGCAACAGACAGTACGCC 356
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 spValGlnValIleuAspIlySerGlnIlnGluMetIleuPheAsnArg 108
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 CGAGCTCCGCATCG.....TACTATGAACAGTATCATCTACTATAA 397
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 ArgArgGlnIlnArgSerGlnAsnPheAsnPheGlyAlaIlyTrHisThrLeuGlu 125
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
398 TGAATCTATCTTGTGATAGATTAATTAAGTGAAGGATCCGATATGC 447
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 uGlnIlnSerGlnIlnMetAspAsnIlnLeuAlaIlnHisProGlyLeu 142
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 TTCAAAATATCCACATGATGCTCTATTTGAGAAGTACCCAGCTATAGT 497
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 alSerIlyValAsnIlnIleGlySerSerPheGlnAsnIlnArgProMetAsnVal 158
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
498 TTAAAGGTTTCT...GCAAAAGAACAAACAGCCAAATATCATAGAT 544
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159 LeuIlySerPheSerThrGlyGlyAsp.....LysProAlaIleTrpIle 172
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545 TGAGCTGGAAATCCATGCGCAGAAATGATGATCTCTCGCTTCTGCTGT 594
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 uAspAlaGlyIlnHisIlnAlaArgIlnTrpValIlnGlnAlaThrAlaLeu 189
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595 GGTTCATAGGC.....GCAAAAGAACAAACAGCCAAATATCATAGAT 605
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 rPheIlnAsnIlyIleValSerAspIlyGlyIlyAspProSerIleThr 205
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 ..... 605
206 SerIleLeuAspAlaLeuAspIlePheLeuProValIlnHisAsnProAs 222
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
606 .....CATATGCAATGTGGAGAAACAAACGCTT 633
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 pGlyTyIValPheSerGlnThrIlyAsnArgMetIlnPArgIlyIlnHis 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
634 CTTTCTATCGCAACATTCATGTCGAGCAACAGCTGTAATGCAACTT 683
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 eIlyValSerGlySerLeuGlyValIglYAlaAspProAsnArgAsnTrp 255
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
684 GTCCTCAACACACGCTGTGAGCAAGTGCATGCTCTCTCATGCTCGGA 733
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 ...AspAlaGlyPheGlyIlyProGlyAlaSerSerAsnProCysSerAs 271
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
734 AACCTCTGTGCTTATCTGATCGACAGCAAGCAAGGAGGAGCTGG 783
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 pSerTyIlnHisGlyProSerAlaAsnSerIlnValIglValIlySerIle 288
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
784 CTAGTTCTTGAGAAATATCAACAGATTAAGCATATCATCAGCATG 833
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

288 alaPheIleIlySerHis...GlyIlyValIlyAlaIlnIleLeu 303
834 CATTCATATCTCCAGCATATAGTGTTCATATTCATATTCATACAGAACTAA 883
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 HisSerTyIlnSerGlnIlnLeuMetCpPheProTyIlnIlyTyIlnCysTrpIly 320
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
884 AAGCAAGACCATGAGAGACTGTCTAGTACGACAGTGAAGCAAGTTGCTG 933
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
320 sLeuAspPheAspIlnLeuSerGlnValAlaGlnIlyAlaIlnGlns 337
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
934 CTATTCGCAAAACATAGTAAAAATCCAGGTATATACATGCGCATGGCTCA 983
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 eIleu...ArgSerLeuHisGlyIlnIlyTyIlnIlyValIglProIleCys 352
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
984 GAAACCTTATATCATGCTCTGAGAGTGGAGAGATGATGATATGATTT 1033
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 SerValIleTyIlnGlnAlaSerGlyGlySerIleAspIlnSerTyIlnAsp 369
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1034 GGGCATCAAAATATTCGTTTACATCAAC..... 1061
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
369 rGlyIlnIleTyIlnSerPheAlaPheGlnIlnLeuArgAspThrGlyArgTyIln 386
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1062 .....CCACCTGTAGAGACGTTTGGCG 1085
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 IyPheLeuLeuProAlaArgGlnIlnLeuPro 396
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: pIrl:S29127
seq_documentation_block:
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
M:Alternate names: pancreatic carboxypeptidase A1
C:Species: Homo sapiens (man)
C:Date: 25-Feb-1994 #sequence revision 19-Jan-1996 #text change 18-Jun-1999
C:Accession: S29127; A34205; S08253; S02810; S71394; S02811
R:Catsum, L.; Villagas, V.; Pascual, R.; Aviles, F.X.; Wicker-Piquant, C.; Puigser
Biochem. J. 287, 299-303, 1992
A:Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A
A:Reference number: S29127; MUID:93038569
A:Accession: S29127
A:Molecule type: mRNA
A:Residues: 1-419 <CAT>
A:Cross-references: EMBL:X67318; NID:935329; PIDN:CAA47732.1; PID:935330
R:Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
A:Title: Human carboxypeptidase A identifies a BglII RFLP and maps to 7q31-qter.
A:Reference number: A34205; MUID:90196012
A:Accession: A34205
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 330-396 <STE>
A>Note: The authors translated the codon CTG for residue 391 as Val
R:Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A:Title: Further studies on the human pancreatic binary complexes involving procarbox
A:Reference number: S08253; MUID:90169111
A:Accession: S08253
A:Molecule type: protein
A:Residues: 17-43; 'xxx', 114-135 <MOU>
R:Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A:Title: Purification and properties of five different forms of human procarboxypepti
A:Reference number: S02809; MUID:89153096
A:Accession: S02810
A:Molecule type: protein
A:Residues: 17-42 <PAS>
R:laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Wal
Arch. Biochem. Biophys. 332, 8-18, 1996
A:Title: Expression and characterization of human pancreatic preprocarboxypeptidase A
A:Reference number: S71394; MUID:96400327
A:Accession: S71394
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-419 <LAE>

```

C:Genetics:
 A:Gene: GDB:CPA1: CPA
 A:Cross-references: GDB:120597; OMIM:114850
 A:Map position: 7q32-7qter
 C:Superfamily: carboxypeptidase
 C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:117-110/Domain: activation peptide #status predicted <ACP>
 F:111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
 F:179,182,306/Binding site: zinc (His, Glu, His) #status predicted
 F:248-271/Disulfide bonds: #status predicted
 F:358,380/Active site: Tyr, Glu #status predicted

alignment_scores:
 Quality: 489.50 Length: 401
 Ratio: 2.128 Gaps: 13
 Percent Similarity: 57.357 Percent Identity: 32.668

alignment_block:
 US-09-980-881-1 x S29127 ..

Align seg 1/1 to: S29127 from: 1 to: 419

```

24 CTTTGACGCTTCGACGCTTCCTTACCATTTCTCTTCTGTGACGACGA 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 LeuValValLeuSerValLeuValLeuGlyAlaValPheGlyLysGluAsp.. 19
74 TGTCTTCGCGCTTCGACGAGTGCGCAAGTTCTAGCTGTCTTCTTGAACCT 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20 .....PheValGlyHisGlnValLeuValArgLysSerValAlaAspG 33
124 CTAGCAAGTTCAGATTCTACAGAACTTACTACACATATGAGATTGTT 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
33 LuValGlnValGlnLysValLysGlnLysGluAspLeuGlnHisLeuGln 49
174 CTC.....TGGCAG.....CCGTTAAACGCTGA 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
50 LeuAspPheTrpArgGlyProAlaHisProGlySerProIleAspVal.. 65
197 CCTATTGTGAAAGAAACAGTCCATTTTGTGTAATGATCATCTGTG 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66 .....ArgValProPhe.....ProSerI 72
247 TCGACAATGTAAGAGCCATTAAATGTAGACGGAATTCATGACGTGC 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
72 LeuGlnValValLysIlePheLeuGlnSerHisGlyIleSerTyrGlnThr 88
297 TTGCTGCACAGCTGGAAGATCTTAT.....CAACAGCAGATT.. 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 MetIleGlnAspValGlnSerLeuAspGlnGlnGlnIleMetLph 105
336 .....TCCACGACACAGTCAGCCCGAGCT 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 eAlaPheArgSerValArgSerThrAspThrPheAsn..... 118
364 CCGCATGCTACTATGAACAGTACACTCACTAAATGAATCTATCTTGG 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 .....TyrAlaThrTyrHisThrLeuGlnGlnIleTyrAspPhe 131
414 ATGAATTTTAACTGAGAGCATCCGATATGCTTACAAAAATCCACAT 463
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 LeuAspLeuValAlaGlnAspProHisLeuValSerLysIleGlnI 148
464 TGAATCCATTTGAGAGTACCACTCTATGTTTAAAGTTTCTGGA 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 eGlyAsnThrTyrGlnLysArgProIleTyrValLeuLysSer...T 164
514 AAGAACAACAGCCAAAATGCATATGATGATGATGATGATGATGCC 563
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
164 hrgLysLysSerLysArgProAlaIleTrpIleAspThrGlyIleHisSer 180
564 AGAAGATGATCTCTCGCTTCTGCTTGCTTGCTGCTGCTGCTGCT 602

```

```

181 ArgGlnTrpValThrGlnAlaSerGlyValTrpPheAlaLysLysIleThr 197
602 .....
197 rGlnAspTyrGlnLysAspAlaAlaPheThrAlaIleLeuAspThrLeuA 214
602 .....
214 spIlePheLeuGlnIleValThrAsnProAspGlyPheAlaPheThrHis 230
603 GGCCATATGCAATGTGAGAAAGAACCGTCTTCTTCTGGAACAACGA 652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 SerThrAsnArgMetTrpArgLysThrArgSerHisThrAlaGlySerLe 247
653 TTGCATCGAACAAGACCTGATACCACTTGTCTCCAAACACTGTGTG 702
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
247 uGlyIleGlyValAspProAsnArgAsnTrp...AspAlaGlyPheGlyL 263
703 AGAAGTGCATCCAGTTCCTCATGCTCGAAGACCTGATGACTTTAT 752
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
263 euserGlyAlaSerSerAsnProCysSerGlnThrTyrHisGlyLysPhe 279
753 CCTGAGTGCAGAACGAGATGAGGCACTGGCTACTTCTTGAGAGAA 802
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 AlaAsnSerGlnValGlnValLysSerIleValAspPheValLysAspH 296
803 TATCAACCAAGATTAAAGCATCATCAGCATGATTCATCCGACGATA 852
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
296 sGlyAsn...IleLysAlaPheIleSerIleHisSerTyrSerGlnLeu 312
853 TAGTGTTCATATTCCTATACAGCAAGTAAAGCAAGACCATGAGGAA 902
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 eumetTyrProTyrGlyTyrLysThrGluProValProAspGlnAspL 328
903 CTGTCCTCTAGTACGACGAGTGTGCTGCTATGCAAACTAGTAA 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 LeuAspGlnLeuSerLysAlaIleValThrAlaLeu...AlaSerLeuT 344
953 AAATACAGATATACATACATGCGCATGCTCAGAAACCTTATACCTG 1002
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
344 rGlyThrLysPheAsnTyrGlySerIleIleLysAlaIleTyrGlnAla 361
361 eArgLysSerThrIleAspTrpThrTyrSerGlnGlyIleLysTyrSerPhe 377
1053 ACA 1055
   |||
378 Thr 378

```

seq_name: p1r2:T33527

seq_documentation_block:
 hypothetical protein T06A4.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33527
 R:WU, X.; Tio-Wollam, A.; Ozersky, P.; Wilson, R.
 submitted to the EMBL data library, October 1998.
 A:Description: The sequence of C. elegans cosmid T06A4.
 A:Reference number: Z21364
 A:Accession: T33527
 A:Status: preliminary; translated from GB/EMBL/DBU
 A:Molecule type: DNA
 A:Residues: 1-528 <WU>
 A:Cross-references: EMBL:AF098994, PIDN:AA67473.1, GSPDB:GN00019, CESP:T06A4.1
 A:Experimental source: strain Bristol N2; clone T06A4
 C:Genetics:
 A:Gene: CESP:T06A4.1
 A:Map position: 1
 A:Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2

alignment_scores:

Quality: 437.50 Length: 424
Ratio: 1.878 Gaps: 16
Percent Similarity: 54.953 Percent Identity: 30.425

alignment_block:

US-09-980-881-1 x T33527 ..

Align seg 1/1 to: T33527 from: 1 to: 528

```

18 ATGAGCTTTCAGC...CTGCAGCTCCTTACCACATGCTCTCTCTG 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 MethTrpScysGlnIleLeuAlaSerLeuThrSerPheTrpVa 17
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 TGACGACCATGCTCTCCGCTTCAGAGTGC..... 95
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 LpheArgHisValLeuAlaThrThrAspGlyArgAlaPheArgLys 34
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
96 ..... 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 eSerAlaThrThrProPheAsnGlnLysThrSerSerPheLysIle 50
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 GCTGCTCTTCCTAGACCTCTAGCAGTTCAGTTCAGTTCAGATCTT.. 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 ArgLysAsnProGlnThrGlnGlySerValLysTrpLeuArgSerLeu 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 .....ACTACACATATAGATTTCTCTGCAGCCGCTACAC... 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 rGluAspSerSerProGlyIleLeuAspPheTrpGlnProProthrasn 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192 .....GCTGACCTATCTGTGAAGAAAAACAAGTCATTTT 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 legLysAlaIleValAspLeuThrValAlaProAlaAspAlaProAlaPhe 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 GTAATTCATCTGTATGTGCACAAATGTGAAGCCCATTAATGTGACGCG 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 Val.....LysAspLeuGlnSerLysLys..... 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 AATTCCATGCAAGTCTCTGCTGCAGACGTGAAGATCTTATTCACAGC 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109 .IleSerTrpIleValAlaValAsnAspLeuSerLysAlaIleGlnAsnG 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 AGATTTCAGACGACACAGTCCGACCCCGACGCTCCGACGTACTAGAA 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 LuArgGlySerAspLysPheTrpAsnProValAlaGlyPheAlaTrpAsp 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
381 CAGTATCACTACTAAATGAAATCTATCTTGGATAGATTAATTAATCTGA 430
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 LysTrpAsnSerLeuGlnGlnIleGlnThrGlnMetLysArgLeuLysI 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 GAGGCATCCGATATGCTTACAAAATCCACATGGATGCTCATTTGAGA 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
158 ArgLysTrpThrMetIleThrLeuIleAspIleArgIleGlnSerHisGln 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
481 AGTACCACACTATGTTTAAAGTTTCGAAAAGAACAAADA..... 524
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 snArgThrLeuLeuValMetLysIleThrGlyLysTrpAsnProLeuGly 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 GCCAAAATGCGCATATGATGATGATGATGATGATGATGATGATGATGAT 574
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192 SerLysIleSerMetTrpIleAspAlaGlyIleHisAlaArgIleLysTrp 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
575 CMTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
208 eAlaProAlaThrAlaMet...TyrIleAlaHisGlnLeuLeuGlyT 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
608 ..... 608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 yrcLusAsnAspAlaThrValAlaLysLeuMetAspHisIleAspPheTr 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
609 .....AA 610
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

241 IleuProValMetAsnProAspGlyTrpGlnLysTrpArgGlnLysAs 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
611 TCGAATGTGGAGAAAGAACGCTCT.....TTCT 639
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257 nArgMetTrpArgLysAsnArgSerProAlaLysCysAlaArgGlnThr 274
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
640 ATGCGAACATCATTCATGATCGACAGACAGCCGATATACCATTTGCTCC 689
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 heserThrValCysCysSerGlyValAspLeuAsnArgAsnPhe...Asp 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
690 AAACACTGGTGTGAGAAAGTGCATCCAGTTCTCATGCTCGGAACCTA 739
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
290 TrpPheTrpAlaSerThrGlySerSerAspProCysHisAspThrTy 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
740 CTGTGACCTTATTCCTGATGACAGACAGAAAGTGAAGCGAGTGGTAGT 789
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306 rHisGlySerAlaAlaPheSerGlnProGlnSerGlnAlaValArgAsp 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
790 TCTTGAGAAAGAAATATCACACGATTAAGCATATACATGATGATGATCA 839
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323 heLeuGlnGlnAsnThrProGlu.....AlaPheIleSerLeuHisSer 337
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
840 TACTCCGACATATAGTGTTCATATTCATATTCATACAGA...AGTAAAG 886
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 TyrSerGlnMetTrpLeuIleProTyGlnHisArgLysGlnSerTrp 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
887 CAAAGAC...CATGAGGACCTGCTCTAGTACGACGAGTGCAGTTCGCTG 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 ogLysAspTrpHisThrGlyLeuArgProLeuAlaLeuAlaThrLysA 371
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
934 CTATTGACAAAACCTAGTAAATAATCCAGTATACATGCGCATGCGCTCA 983
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 leuLysTrpGlnLeu...TyrGlyThrLysTrpGlnValGlyThrGlyAla 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
984 GAAACCTTATCCATGCTCTGAGGAGTGGGAGCATGATGATCAT...CA 1030
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 AspLeuMetTrpGlnLysSerGlyGlySerHisAspTrpAlaLysGly 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1031 TTTGGGCATCAAAATATTCGTTT 1052
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 nLeuLysValProTyAlaLeuTr 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: p1r2:T33526

seq_documentation_block:
hypothetical protein T06A4.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33526
R:Wu, X.; Yin-Mollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of *C. elegans* cosmid T06A4.
A:Reference number: Z1364
A:Accession: T33526
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-667 <WU>
A:Cross-references: EMBL:AF098994; PIDN:AA67474.1; GSPDB:GN00019; CESP:T06A4.3
C:Genetics:
A:Gene: CESP:T06A4.3
A:Map position: 1
A:Introns: 50/3; 96/2; 167/3; 236/3; 276/3; 315/1; 365/3; 385/3; 444/1; 487/1; 582/3

alignment_scores:

Quality: 400.50 Length: 483
Ratio: 1.655 Gaps: 17
Percent Similarity: 50.104 Percent Identity: 27.122

alignment_block:

US-09-980-881-1 x T33526 ..

Ratio: 2.644 Gaps: 8
Percent Similarity: 62.768 Percent Identity: 37.709

Alignment_block:
US-09-980-881-1 x CBPB_RAT

Align seg 1/1 to: CBPB_RAT from: 1 to: 415

```

18  ATGAGCTTTGAGCGCTTGAGTCCCTGATCCCATGTTCTCTCTGTA 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1  MetLeuLeuLeuLeuValSerValAlaLeuAlaHisAlaSerGI 17
68  GCAGCATGTCTTGGCGTCCAGAGTGCCCAAGTTCAAGTCTCTCTTA 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17  uGIuHIS.....PheaspGlyasnArgValITyArgValSerValH 31
118  GAACCTTAGGCAAGTTCAAGTCTTACTACAGAACTTACACATATGAG 167
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
31  IselGlyIuaspHisValasnLeuIleGlnGluLeuAlaasnThrLysGIu 47
168  ATTGTTCTTGAGCGCGCGGTACAGCTGACCTTATGTGAGAAACA 217
   ||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
48  IleaspPheThrLysProaspSerAlaThrGlnValLysProLeuThr 64
218  AGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTGAAGCCCAT 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64  rValaspPheHisValLysAlaGluaspValAlaaspValGluaspPhe 81
268  TAAATGTGAGCGGAATTCATGCGATGCTGTCTGGCGAGCGTGAAGAT 317
   ||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
81  euGIuIuaspnGluValHisTyrGlnValLeuIleSerasnValArgasn 97
318  CTTATTCACAGCAGATTTCCACAGACAGACAGTACGCCCGGACCTCCG 367
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
98  AlaLeuGlnSerGlnPheaspSerHisThr.....ArgAlaSerGI 111
368  ATCGTACTATGAACAGTATCAGTCACTCAATGAATCTATCTTGATAG 417
   : ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
111  yHisSerTyrThrLysTyrAsnLysTyrLysThrIleGluAlaThrIleG 128
418  AATTATTAATCTGAGAGCATCTGATATGCTTCAAAAATCCATATGGA 467
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
128  InGIuValAlaThrAspAsnProaspLeuValThrGlnSerValIleGIy 144
468  TCCATCTTTGAGAAATCCCACTCTATGTTTAAAGTTCTTGGAAGA 517
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
145  ThrThrPheGlnGlyArgAsnMetTyrValLeuLysIle...GlyLys 160
518  ACAAAACAGCCAAAATGCATATGATGATGACGTGTGAATCCATCCAG 567
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
160  rArgProaspnLysProAlaIlePheHisLeaspCysGlyPheHisAla 177
568  AATGATCTCTCTGCTTCTCTGCTTGGTTCATA..... 602
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
177  LutrPleaserProAlaPheCysGlnTyrPheValArgGluAlaValArg 193
602  ..... 602
194  ThrTyrAsnGlnGlnIleHisMetLysGlnLeuLeuaspGluLeuasp 210
603  .....GGCCAT..... 608
210  eTyrValLeuProValValasnIleLeaspLysValTyrThrThrThr 227
609  ..AATGATGTGAGAAAGAACCGTTCTTCTATGCAACAATCAATTCG 656
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
227  yAspArgMetThrParGlyThrArgSerThrMetLacIleSerSerCys 243
657  ATCGGAACAGACCTGAAATGCAATTTGTCTCCAAACACTGCTGAGGA 706
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
244  LeuGlyValAlaArgProAsnArgAsnPhe...AsnAlaGlyTyrCys 259
707  AGGTGCATCAGTCTCTCATGTGCGAAACCTACTGTGACCTTTATCCTG 756

```

```

259  IGIyAlaSerArgSerProCysSerGIuThrTyrCysGlyProAlaProG 276
757  AGTCAGAAACAGAGAGGAGGAGCTAGTTCTTGAAGAAATATC 806
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276  IuSerGIuLysGIuThrLysAlaLeuAlaaspPheIleArgAsnLeu 292
807  AACCGATTAAGCATATCATCAGCATGATTCATCTCCAGCATATAGT 856
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
293  SerThrIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIle 309
857  GTTCCATTTCTATACACAGTAATAAGCAAGACATGAGGAAGTGT 906
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
309  uTyrProTyrSerTyrAspTyrLysLeuProGluAsnTyrGluGluLeu 326
907  CTCTAGTAGCAGGAGCAGTTCGTCTTGTGCAAAATCTATTAATAAT 956
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
326  snAlaLeuValLysGlyAlaAlaLysGluLeu...AlaThrLeuHisGIy 341
957  ACCAGGTATACACATGGCGCATGGCTCGAGAAACCTTATACCTGCTG 1006
   ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
342  ThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrProAlaAlaGI 358
1007  AGGTGGGAGCAGATTGCATCTATGATTTGGCATCAAAATATGCTTACAT 1055
   ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
358  yGIuSerAspAspTyrSerTyrAspGlnGlyIleLysTyrSerPheThr 375
1056  C..... 1057
375  heGIuLeuAlaArgAspThrGlyPhePheGlyPheLeuLeuProGluSerGln 391
1058  ...AAACCCACCTGTAGAGAGAGCTTTGCCGCTGTCTTAAATAGCTTG 1104
   :: :: :: :: :: :: :: :: :: :: :: :: :: ::
392  IleArgGlnThrCysGlnGluThrMetLeuAlaValLysTyrIleAlaAs 408
1105  GCATGTC 1111
408  nTyrVal 410

```

seq_name: SwissProt_40:CBPC_MOUSE

seq_documentation_block:
ID CBPC_MOUSE STANDARD; PRT; 417 AA.

```

AC P15089;
DT 01-APR-1990 (Rel. 14, Created)
DI 01-APR-1990 (Rel. 14, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Serafin W.E.;
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT novel member of the carboxypeptidase gene family.";
RL J. Biol. Chem. 264:20094-20099(1989).
CC -1 CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1 SUBCELLULAR LOCATION: Secretory granules.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

[illegible]

480	AACTACCACCTCATGTTTAAAGGTTTCGGAAAAGAACAAACAGCCA	523
151	AspAspIleuLeuYrValLeuIlys..GIlyIlysAspIgluAr	166
530	AAATGCCATATGATTCAGTGTGGAATCCATGCCAGATGATTCCTC	579
166	gluysAlaIlePheMetAspCysIlyIleHsAlaIatrgIuITrIleSer	183
580	CTGCTTTTGTCTTGTGTTCA..GGCAT	608
183	roAlaPheCysIuITrPheValYrGlnAlaIatrgIysSerYrGlyLys	199
609	AAAT.....	611
200	AsnIysIleMetIhrLysLeuAspArgMetAsnIheYrValLeuPr	216
612CGAATGT	618
216	oValrPheAsnValAspIglYrIleITrSerITrPthGlnAspArgMet	233
619	GGAGAAAGAACCGTCTTCTTATGCGAACATCATTCATCGGAACAC	668
233	rPrArgIysAsnArgSerArgAsnGlnAsnSerITrCysIleGlyThrAsp	249
669	CTGAATAGCACTTTGTCTCCAAACACTGGTGTGAGAGATGCATCCAG	718
250	LeuAsnArgAsnIhe..AspValSerITrAspSerITrProAsnITrAs	265
719	TTCCATGCTCGGAACACTGTGGACTTTATCCGAGTCAGAACACAG	768
265	nIysProCysLeuAsnValYrArgIglYrProAlaProGlnSerIuIysG	282
769	AAATGAAGCAGATGGCTAGTTCTTGAGAAATATCAACACAGATTAA	818
282	IuIhrLysAlaValThrAsnPheIleArgSerHsIleuAsnSerIleLys	298
819	GCATATATGACATGCATTCATCTACCTCCACATATAGTGTTCATATTC	868
299	AlatYrIleIhrPheHsISerYrSerGlnMetLeuLeuIleProYrGly	315
869	CTATACACGACAGTAAGCAAAAGACATCAGAGAACTGTCTAGTACCA	918
315	YrYrITrIhrPheLysLeuProIroPsnHsGlnAspIleuLysValAla	332
919	GTGACACAGTTCGTCTATTCACAAAATCTGTAAATATCCAGGTATACA	968
332	rgIleAlaITrAspAlaLeu...SerITrArgYrGluITrArgYrIle	347
969	CATGGCAGTCGCACAAACCTTATACCTAGCTCCCTGGAGTGGGAGCA	1018
348	YrGlyIProIleAserITrIleYrLysITrSerIySerSerLeuAs	364
1019	TTGGATCATGATTTGGCGCATCAATATTCGTTAC.....	1054
364	PrTrValITrYrAspIleuGlyIleLysHsITrPheAlaPheGluLeuArg	381
1055ATCAATCCACCC	1066
381	sPrLysGlySerGlyPheLeuLeuProGlnSerArgIleLysProITr	397
1067	TTGAGAGAGAGCTTTGGCGCTGTCGTAAATAGCTGGCATGTCATTA	1116
398	CysLysGluITrMetLeuSerValLysPheIleAlaLysITrIleLeuLys	414
1117	GAAT	1120
414	sAsn	415

AC p15086; 060834; 01-Apr-1990 (Rel. 14, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (pancreas-specific protein) (PASP)
 GN CPB1 OR CPB OR PCPB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
 RC TISSUE=Pancreas;
 RX MEDLINE=92129345; PubMed=1370825;
 RA Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S., French C.K.;
 RT "Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B.";
 RL J. Biol. Chem. 267:2575-2581(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=99182241; PubMed=9524066;
 RA Alay P., Catasus L., Vallegas V., Reverter D., Vendrell J., Aviles F.X.;
 RT "Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B.";
 RL Biol. Chem. 379:149-155(1998).
 RN [3]
 RP SEQUENCE OF 16-43.
 RC TISSUE=Pancreas;
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT "Purification and properties of five different forms of human procarboxypeptidases.";
 RL Eur. J. Biochem. 179:609-616(1989).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O = peptide + L-lysine(or L-arginine).
 CC -1- TISSUE SPECIFICITY: PANCREAS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE ZINC CARBOXYPEPTIDASE FAMILY
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M81057; AAA66973.1; -
 DR EMBL: AJ224866; CAA12163.1; -
 DR PIR: S02812; S02812.
 DR PIR: A42337; A42332.
 DR HSP: P09955; INSA.
 DR MEROPS: M14.003; -
 DR MIM: 114852; -
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; zn_carboxept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; zn_carboxept. 1.
 DR PRINTS: PR00765; CRBOXPTASA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 110 ACTIVATION PEPTIDE.
 FT CHAIN 111 417 CARBOXYPEPTIDASE B.
 FT METAL 176 179 ZINC (BY SIMILARITY).
 FT METAL 179 179 ZINC (BY SIMILARITY).
 FT METAL 304 304 ZINC (BY SIMILARITY).

FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 173 186 BY SIMILARITY.
 FT DISULFID 245 268 BY SIMILARITY.
 FT DISULFID 259 273 BY SIMILARITY.
 FT CONFLICT 16 16 H -> A (IN REF. 1; AA SEQUENCE).
 FT CONFLICT 17 17 H -> Q (IN REF. 1; AA SEQUENCE).
 FT CONFLICT 37 37 H -> Q (IN REF. 3).
 FT CONFLICT 208 208 N -> D (IN REF. 1).
 FT CONFLICT 245 245 MISSING (IN REF. 1).
 SQ SEQUENCE 417 AA; 47366 MW; BB1CF212D830305E CRC64;
 Alignment scores:
 Quality: 653.50 Length: 422
 Ratio: 2.394 Gaps: 8
 Percent Similarity: 64.692 Percent Identity: 33.649
 alignment_block:
 US-09-980-881-1 x CBPB_HUMAN ..
 Align seg 1/1 to: CBPB_HUMAN from: 1 to: 417
 24 CTTTCAGCCCTGACGCTGTCACCATGCTGCTGTCGAGCGCA 73
 1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisI 17
 74 TGTCTTCGCG...TTCCAGAGTGGCCAGTCTAGCTGCTCTCTAGAA 120
 17 sglYglYgluHisPhegluYgluYValPheArgValAlaValAla 34
 121 CCTCTAGCGCAGTTCAGTTCATGAGATCTTACTGATCAATGATGAT 170
 34 spgluSnHisIleAsnIleLeuArgIleuAlaSerThrThrIle 50
 171 GTTCTGGAGCGCGGTACAGCTGACCTTATGTGTAAGAAACAGACT 220
 51 AspPheThrIlyPProAspSerValThrGluIleYProHisSerThyA 67
 221 CCATTTTGTGTAATGATCATGTGTCGACAAATGTAAGCCATTGA 270
 67 lAspPheArgValValYsAlaGluAspPheValThrValGluAsnValLeu 84
 271 ATGTGAGCGGAATTCATGATGATGAGTCTGTCGTCGACAGTGAAGAT 320
 84 ysluInsngluLeuGlnIlyIlyValLeuIleSerInsLeuArgAsnVal 100
 321 ATTCACAGCAGATTTCACAGACAGACAGTACGCCCGAGCTCCGATC 370
 101 ValGluAlaGlnPheAspSerArgVal.....ArgAlaThrGluY 114
 371 GTACTATGAAAGTATCATCTACTCAATGAATGAAATCTTCTGGATGAAT 420
 114 sSerIlyGluIlyIlyTrAsnIlyStrpGluThrIleGluAlaThrPheGln 131
 421 TTATTAAGTAGAGCAGATCCATGATATGCTTACAAATAATCCATGATGCC 470
 131 lValAlaThrGluAsnProAlaLeuIleSerArgSerValIleGluY 147
 471 TCATTGAGAGTAACCATCTATGTTTAAAGTTTCTGGAAAAGAAC 520
 148 ThrPheGluIlyArgAlaIleIlyLeuLeuYsVal...GlyIysAlaG 163
 521 AACAGCCAAATAGCCATATGATGATGATGATGATGATGATGATGAT 570
 163 ygluAsnIlyProAlaIlePheMetAspCysglYpPheHisAlaArgGlu 180
 571 GGAATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
 180 rPheSerProAlaPheCysglIlyThrPheValArgGluAlaValArgThr 196
 602 602
 197 TyrlYArgGluIleGlnValThrGluLeuAsnIlyLeuAspPheY 213

```

603 .....GGCCAT.....A 609
213 rValleuProValleuAsnIleAspGlyTyrIleTyrThrPrThyrLys 230
610 ATGCAAGTGGAGAAAGAACGCTTTCTTCATGGCAACAATCATTCGCTC 659
230 eraArgpHePrArgLysThrArgSerThrHisThrGlySerSerCysIle 246
660 GGAACAGACCTGATATACCACTTGTCTCCAAACACTGGTGTGAGAGAG 709
247 GlyThrAspProAsnArgAsnPhe...AspAlaGlyTrpCysGluIleG1 262
710 TGCATCCAGTTCCTCATGCTCGGAACCTACTGTGACCTTATCCGTAGT 759
262 yAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaIleGus 279
760 CAGACACGAGAGTGAAGCACTGGCTAGTCTTCTTGAGACAGAAATATAC 809
279 ergLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295
810 CAGATTAAGCATATCATGATCATATCATATCCAGCATATATAGTGT 859
296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleTyr 312
860 TCCATATTCCTATACACGAGTAAACCAAGACCATGAGCACTGTCTC 909
312 rProTyrSerTyrAlaTyrLysLeuGlyLysAsnAlaIleLeuAsnA 329
910 TAGACGCCAGTGAAGCACTGCTCTATTCGACAAACCTAGTAAATATCC 959
329 lalaLeuAlaLysAlaThrValLysGluLeu...AlaSerLeuHisGlyThr 344
960 AGGATATACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGGAGG 1009
345 LysTyrThrTyrGlyProGlyAlaThrThrIleTyrProAlaIleGlyG1 361
1010 TGGGAGCAGTATGATTCATGATTTGGCCATCAAAATATTCGTTTACATC. 1057
361 ySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPheG 378
1057 ..... 1057
378 luleuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394
1058 AAACCCACTGTAGAGAGCTTTTGCCTGCTCTTAATAGCTTGCGCA 1107
395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaIleSerTyr 411
1108 TGTGATTAGCAAGTGT 1123
411 rValleuGluHisLeu 416

seq_name: SwissProt_40:CBPC_HUMAN
seq_documentation_block:
ID CBPC_HUMAN STANDARD: PRT; 417 AA.
AC P15088;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=90083291; PubMed=2594780;
RA Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,

```

```

RA Austen K.F., Serafin W.E.;
RT "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT comparison of the protein with mouse mast cell carboxypeptidase A and
RT rat pancreatic carboxypeptidases."
RT Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mast cells;
RX MEDLINE=92105393; PubMed=1729276;
RA Reynolds D.S., Gurley D.S., Austen K.F.;
RT "Cloning and characterization of the novel gene for mast cell
RT carboxypeptidase A."
RL J. Clin. Invest. 89:273-282(1992).
RN [3]
RP SEQUENCE OF 110-417 FROM N.A.
RX MEDLINE=92333165; PubMed=1629626;
RA Natsunaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsunaki M.,
RA Wintroub B.O., Rutter W.D., Goldstein S.M.;
RT "Human skin mast cell carboxypeptidase: functional characterization,
RT cDNA cloning, and genealogy."
RL J. Invest. Dermatol. 99:138-145(1992).
RN [4]
RP SEQUENCE OF 110-137.
RX MEDLINE=69214692; PubMed=2708524;
RA Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.O.;
RT "Human mast cell carboxypeptidase. Purification and
RT characterization."
RL J. Clin. Invest. 83:1630-1636(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M27717; AAA35652.1; -.
DR EMBL: M73720; AAA59568.1; -.
DR EMBL: M73716; AAA59568.1; JOINED.
DR EMBL: M73717; AAA59568.1; JOINED.
DR EMBL: M73718; AAA59568.1; JOINED.
DR EMBL: M73719; AAA59568.1; JOINED.
DR EMBL: S40234; AAB22578.2; ALT-SEQ.
DR PIR: A43929; A43929.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR MTM: 114851; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_Carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_Carboxypept. 1.
DR PRINTS: PR00765; CRBOXYPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109
FT CHAIN 110 417
FT METAL 176 176
FT METAL 179 179
FT METAL 304 304
FT ACT_SITE 378 378
FT DISULFID 173 186
FT FT 245 268
SO SEQUENCE 417 AA; 48700 MW; 8CB90DB758117B24 CRC64;

```

alignment_scores:

Quality: 647.50 Length: 427
 Ratio: 2.398 Gaps: 10
 Percent Similarity: 63.232 Percent Identity: 33.489

alignment_block:

US-09-980-881-1 x CBPB_HUMAN ..

Align seg 1/1 to: CBPB_HUMAN from: 1 to: 417

```

6  ATTGCTGTTGGATGAGCTTTGACAGCTTGCATCTGACCAATGTT 55
   ::: |||||:::
5  LeuProValGlyLeuIleAlaThrIleAlaIle...AlaProVal.. 19
56 TCTCTCTGTGACGACCATGCTCTGCGCTTCAGAGTGGCCAGTTCTAG 105
   |||||:::
20 ..... ArgPheAspArgGluValAlaPhe 28
106 CTGCTCTTCTAGAACCTGACGAGTTCAGATCTGATGCAATCTTACT 155
   ::: |||||:::
28  rglValIysProGlnAspGluLysGlnAlaAspIleIleLysAspLeuAla 44
156 ACAACATATGAGATTGTTCTGCGACGCGGTAAACAGCTGACCTTATGT 205
   |||||:::
45  LysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisIleValAl 61
206 GAAGAAAAACAAGTCCATTTTGTAAATGCACTGATGTCGACAAATG 255
   : :::: |||||:::
61  AlaAsnMetMetValAspPheArgValSerGluLysGlnSerGlnAlaI 78
256 TGAACCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTGCTGGCA 305
   ::::: |||||:::
78  LeuIleSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94
306 GAGCTGGAAGATCTTATTCACAGCAGATTTCC...AAGCACAGCTTCAG 352
   |||||:::
95  AspLeuGlnGluGluIleGluLysGlnPheAspValLysGlnSplLepr 111
353 CCCCCGACCTCCGCGATGCTACTATGACACATATCACTCAATATGAAA 402
   |||||:::
111  OGlyArgHisSer.....TyrAlaLysTyrAsnAsnTrpGluLysI 125
403 TGTATTCCTGATGAATTTATTAATGAGCGATCCTGATATGCTTACA 452
   |||||:::
125  LeuValAlaTrpThrGluLysMetMetAspLysTyrProGluMetValSer 141
453 AAATCCACATTTGATCTCATTTGGAAGTACCACATCTATGTTTAA 502
   ::::: |||||:::
142  ArgIleLysIleGlySerThrValGluAspAsnProLeuTyrAlaLeuL 158
503 GGTTCCTGAAAAAGAACAAACAGCCAAATGCGATATGATGACTGTG 552
   |::: |||||:::
158  sIle...GlyGluLysAsnGluValArgTyrAlaIlePheMetAspCysG 174
553 GAATCCATGCGAGAAATGATCTCTCGCTTCTGCTTGGTGGTCAT 602
   |||||:::
174  LylIleHisAlaArgIleTrpValSerProAlaPheCysGlnTrpPheVal 190
602 .....
191  TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeu 207
603 .....
207  uAspArgMetAsnPheTyrIleLeuProValAlaPheAsnValAspGlyTyrI 224
609 .....
224  letPserTrpThrLysAsnArgMetTrpArgLysAsnArgSerLysAsn 240
642 GCGAACCATATTCATCGAACAGCAGCTGAATAGCAACTTGTCTCCAA 691
   |||||:::
241  GlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe...AsnAl 256

```

```

692  ACACGTGTGAGAGAGTGCATCCAGTTCCTCATGCTCGGAACCTACT 741
   |||:::
256  aserTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrA 273
742  GTGGACTTTTCTGAGTGCAGAACGAGTGAAGGAGTGGCTAGTTTC 791
   |||||:::
273  rglYserAlaProGluSerGluLysGlnValTyrIleThrPheHisSerTyr 289
792  TTGAGAAGAAATATCAACGATTTAAAGCATATGATGATCATATCA 841
   ::::: |||||:::
290  IleArgSerIleAsnAsnGluIleLysValTyrIleThrPheHisSerTyr 306
842  CTCCACGATATATGTTTCCATATTCCTATACAGAAATAAAGCAAG 891
   |||||:::
306  rSerGlnMetLeuLeuPheProTyrGlyTyrThrSerLysLeuProPro 323
892  ACCATGAGAACTGTCTCTAGTACGAGTTCAGTTCGCTATTTGAC 941
   ::::: |||||:::
323  snHisGlnAspLeuAlaLysValAlaLysIleGlyThrAspValLeu... 338
942  AAACATAGTAAATATACAGATATACATGAGCCATGCTCAGAAACCTT 991
   ::::: |||||:::
339  SerThrArgTyrGluTrpArgTyrIleTyrGlyProIleLeuSerThrI 355
992  ATACCTAGCTCTCGAGGTGGGACGATTTGATCTATGTTGGCATCA 1041
   :|||::: |||||:::
355  eTyrProIleSerGlySerLeuAspTrpAlaTyrAspLeuGlyIle 372
1042  AATATTCGTTTAC..... 1054
   |||||:::
372  yshIshThrPheAlaPheGlnLeuAlaArgAspLysGlyLysPheGlyPheLeu 388
1055  .....ATCAACCCACCTGTAGAGACTTTGCGCGCTG 1089
   |||||:::
389  LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaVal 405
1090  CTCTAAATATGCTTGGCATGTCTATAGGAAT 1120
   |||||:::
405  LysPheIleAlaLysTyrIleLeuLysHis 415

```

seq_name: SwissProt_40:CBPB_CANFA

seq_documentation_block:

ID	CBPB_CANFA	STANDARD:	PRT:	416 AA.
AC	P55261;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule membrane associated protein) (ZAP47).			
GN	CPBL.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_TaxID=9615;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	TISSUE=Pancreas;			
RA	Fukuoka S.-I.;			
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =			
CC	peptide + L-lysine(or L-arginine).			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE			
CC	ZINC CARBOXYPEPTIDASE FAMILY.			
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			

```

CC EMBL: D78348; BA11366.1;
DR HSSP; P09955; 1PBA.
DR MEROPS; M14.003;
DR InterPro; IPR003146; Proper_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Proper_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CARBOXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109
FT CHAIN 110 416
FT METAL 175 175
FT METAL 178 178
FT METAL 303 303
FT ACT_SITE 377 377
FT DISULFID 244 267
FT DISULFID 258 272
SQ SEQUENCE 416 AA; 47595 MW; DFE1992CD52F8AB4 CRC64;

```

```

alignment_scores:
  Quality: 640.50      Length: 420
  Ratio: 2.390         Gaps: 9
Percent Similarity: 63.810   Percent Identity: 33.810

```

alignment_block:

US-09-980-881-1 x CBPB_CANPA ..

Align seg 1/1 to: CBPB_CANPA from: 1 to: 416

```

39 GTCCTGATCCACATGTTCTC.....TTCGTGAGCAGCATGT 76
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
5 ILeuValThrLeuAlaLeuAlaSerAlaHisTyrSerGlyLeuHis.. 20
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
77 CTTCGCGTCCAGAGTGGCCAACTTCTAGCTGCTCTCTAGAACTCTA 126
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
21 .....PheGluGlyGlyLeuValPheArgValAsnValGluAspGlu 35
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
127 GGCAGTTCAGTTCTTACAGAACTTCTTACATCATGATGATGTTCTC 176
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
35 snHisLeuAsnLeuHisThrLeuAlaSerThrThrGlnHisLeuAspHe 51
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
177 TGCAGCGCGCTTACAGCTGACCTTATGTTGTAAGAAAACAAGTCACTT 226
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
52 TrpLysProAspSerValThrGlnHisLeuValProHisSerThrAlaAsp 68
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
227 TTTTGTAAATGATCTGATGTCGACAAATGTAAGCCCATTTAAATGTA 276
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
68 eArgValIlyAlaGluAspIleLeuThrValGluAspPheLeuHisGln 85
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
277 GCGGATTCATGACATGCTGTGCTGCGACAGCTGGAAGATCTTATTCAA 326
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
85 snGluLeuHisTyrGluValLeuHisAsnAsnLeuArgLeuValLeuGlu 101
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
327 CACGAGATTTCACAGCAGACAGCTGAGCCCGGAGCTCCGACATGTA 376
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
102 GlyGlnPheGlyArgGlnVal.....ProIleThrGlnHisSerTyr 115
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
377 TGAACAGTATCACTCACTAAATGTAATCTATCTGTGATGAATTTTAA 426
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
115 rGluLysTyrAsnArgTrpGlnHisLeuAlaValTrpThrGlnHisVal 132
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
427 CTGAGAGGATCTCGATATGCTTACAAATCCATGATGATGATCTCTT 476
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
132 HisArgGluAsnProAspLeuIleSerArgTyrSerIleGlyThrThrPhe 148
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
477 GAGAAGTACCCACTATGTTTAAAGTTCTTGGAAAAACAACAACAGC 526
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
149 GluGlyArgThrIleTyrLeuLeuLysVal...GlyLysAlaGlyGlnAs 164

```

```

527 CAAAATGCCATATGATGATGATGATGATGATGATGATGATGATGAT 576
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
164 nLysProAlaIlePheMetAspCysGlyPheHisAlaArgGluTrpLys 181
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
577 CTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
181 eTrpAlaPheTrpGlnTrpPheValArgGlu**IleArgThrTyrGly 197
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
602 ..... 602
198 GluGluHisMetThrGluLeuLeuAspLysLeuAspPheTyrValLe 214
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
603 .....GCCAT.....AATGAA 615
214 uProValGlyAsnHisLeuAspGlyTyrValTyrThrTrpThrLysAsn 231
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
616 TGTCGAGAAAAGAACCGTTCTTCTGATGCGAACAATCATGATCGGA 665
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
231 eTrpArgLysThrArgSerThrGlnValGlyThrAsnGlyValGlyThr 247
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
666 GACCTGATGACAACTTGTCTCCAAACACTGGTGTGAGGAGGATGATC 715
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
248 AspProThrArgAsnPhe...AspAlaGlyTrpCysLysIleGlyAla 263
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
716 CAGTTCCTCATGCTCGGAAACCTAGTGTGATCTTATCTGAGTACAGAC 765
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
263 rArgAsnProCysAspGlnThrTyrCysGlyProAlaAlaGluSerGlu 280
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
766 CAGAGTGAAGGCGATGCTAGTGTCTTGGAGAGAAATATACACCAAT 815
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
280 yGluThrLysAlaLeuAlaAsnPheIleArgSerAsnLeuSerSerIle 296
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
816 AAAGCATATACATGACATGATCTTACCTCCAGCATATGATGTTTCATA 865
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
297 LysAlaTyrLeuThrIleHisSerTyrSerGlnMetLeuTyrProTyr 313
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
866 TTCCTATACAGAAATAAAAGAACATGAGAACTGTCTGTAGTAG 915
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
313 rSerTyrAspTyrLysLeuThrGluAsnAsnAlaGluLeuAsnAlaLeu 330
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
916 CCGATGAGCAGATCGCTGATGATGACAAACTAGTAAATATACAGGAT 965
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
330 lAlaValAlaThrValLysGluLeu...AlaThrLeuHisGlyLysTyr 345
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
966 ACATGATGCGCATGCGCTGAGAAACCTTATACCTGCTGAGGAGGAGA 1015
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
346 ThrTyrGlyProGlyAlaThrThrIleTyrProAlaAlaGlyGlySer 362
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1016 CGATTGATCTATGATTTGGGATCAAAATATCGTTTACATC..... 1057
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
362 rAspTrpAlaTyrAspGlnGlyIleLysTyrSerPheThrPheGluLeu 379
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1058 .....AAACC 1063
379 rGAspLysGlyArgTyrGlyPheAlaLeuProGluSerGlnHisSerPro 395
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1064 ACCTGTGAGAGAGCTTGTGCGCTGTCTTAAATATGATCTGCGATGAT 1113
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
396 ThrCysGluGluThrLeuLeuAlaIleLysHisLeuAlaArgTyrVal 412
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1114 TAGGAATGTT 1123
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
412 uGlnHisLeu 415
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
seq_name: SwissProt_40:CBPB_PIG
seq_documentation_block:
ID CBPB_PIG STANDARD; PRT; 401 AA.
AC P09955;
DT 01-MAR-1989 (rel. 10, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)

```



```

186 laHsMethGluPheLeuAspAsnLeuAspPheTyrValLeuProVal 202
603 .....GGCCAT.....AATCGAATGTGAG 622
203 LeuAsnIleAspGlyTyrIleTyrThrIlePheThrIleAsnAlaMetIlePhe 219
623 AAAGAACCGTCTTCTGATGCGAACATCATGTCGGAACAGACATGTA 672
219 GlyThrArgSerThrAsnIleGlySerSerCysThrGlyThrAspPro 236
673 ATACCACTTGTCTCCAAACACATGCTGTGAGGAAGTGCATCCAGTCC 722
236 snArgAsnPhe...AsnAlaGlyTyrPcysThrValGlyAlaSerValAsn 251
723 TCATGCTCGGAACCTACTGTGACCTTATCTGATGACAGACAGAGT 772
252 ProCysAsnGluThrTyrCysGlySerAlaIleGluSerGluIysGluTh 268
773 GAAGCAGTGGCTAGTCTTGTGAGAGAATATCAACGATTTAAAGCAT 822
268 IlyAlaIleuAlaAspPheIleArgAsnAsnLeuSerSerIleIysAlaT 285
823 ACATCAGCATGCTATCATCTCCAGCATATAGTTCATATTCCTAT 872
285 YIleuThrIleHisSerTyrSerIleMetIleLeuTyrProTyrSerTyr 301
873 ACAGAGATAAAGCAAGAACCATGACATGACATGCTCTGATAGCCAGTA 922
302 AspTyrIysLeuProGluAsnAspAlaGluIleuAsnSerIleAlaIlyG 318
923 AGCAGTCTGCTATGTGACAAACTAGTAAATAATCAGATATACATG 972
318 yAlaValIysGluIleu...AlaSerIleuTyrGlyThrSerTyrSerTyrG 334
973 GCCATGGCTCGAAGACCTATATCTGATGCTGTGAGGGGAGCATG 1022
334 IyProGlySerThrThrIleTyrProAlaIleGlyGlySerAspIleP 350
1023 ATCTATGATTTGGCATCAATAATATTCGTTTACA 1055
351 AlaTyrAsnGlnGlyIleTyrSerPheThr 361

seqname: SwissProt_40:CBPB_BOVIN

seq_documentation_block:
ID CBPB_BOVIN STANDARD; PRT; 306 AA.
AC P00732;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B (Ec 3.4.17.2).
GN CPB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh C.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RA MEDLINE=74260705; PubMed=4833744;
RA Schmidt J.J., Hirs C.H.W.;
RT "Primary structure of bovine carboxypeptidase B. Inferences from the
RT locations of the half-cysteines and identification of the active site
RT arginine.";
RL J. Biol. Chem. 249:3756-3764(1974).
[3]

```

```

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RX MEDLINE=76265065; PubMed=957425;
RA Schmid M.F., Herriott J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RL J. Mol. Biol. 103:175-190(1976).
RN [4]
RP ACTIVE SITE.
RX MEDLINE=70007159; PubMed=5344132;
RA Plummer T.H. Jr.;
RT "Isolation and sequence of peptides at the active center of bovine
RT carboxypeptidase B.";
RL J. Biol. Chem. 244:5246-5253(1969).
RN [5]
RP ACTIVE SITE.
RX MEDLINE=73061487; PubMed=4565668;
RA Kimmel M.T., Plummer T.H. Jr.;
RT "Identification of a glutamic acid at the active center of bovine
RT carboxypeptidase B.";
RL J. Biol. Chem. 247:7864-7869(1972).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-Lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
DR PIR; A00912; CPEOB.
DR PDB; 1CPB; 30-SEP-83.
DR MEROPS; M14.003; -.
DR InterPro; IPR00834; Zn_carboxepr.
DR Pfam; PF00246; Zn_carboxepr; 1.
DR PRINTS; PR00765; CBOXYPTSA.
DR PROSITE; PS00132; CARBOXYPEPT_2N_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_2N_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT METAL 149 163
FT METAL 66 66
FT METAL 69 69
FT METAL 194 194
FT ACT_SITE 246 246
FT ACT_SITE 268 268
FT SEQUENCE 306 AA; 34612 MW; C329D2655C44A172 CRC64;
NUCLEOPHILE.

```

```

alignment_scores:
Quality: 589.50 Length: 304
Ratio: 3.039 Gaps: 6
Percent Similarity: 63.816 Percent Identity: 39.803

```

alignment_block:

US-09-980-881-1 x CBPB_BOVIN ..

Align seg 1/1 to: CBPB_BOVIN from: 1 to: 306

```

375 TATGACAGATACCTACTCAATAAGAAATCTATCTTGATAGAAATTTAT 424
6 TyrGluIysTyrAsnAsnTrpIleThrIleGluAlaIleThrIleGluIleVal 22
425 AACGAGAGGACCTCTGATATGCTTACAAAATTCACATTCGATTCCTCAT 474
22 IAlaSerGluAsnProAspLeuIleSerArgSerAlaIleGlyThrIlePhe 39
475 TTGAGAGTACCACCTATATGTTTAAAGGTTTCGAAAGAACAAACA 524
39 heLeuGlyAsnThrIleTyrLeuLeuIysVal...GlyIysProIlySer 54
525 GCCAAAATGCCATATGATTCATGCTGTGAGATTCATGCCAGAGATGAT 574
55 AsnIysProAlaValAlaPheMetAspCysGlyPheHisAlaArgGluTrpI 71
575 CTCCTCGCTTCTGCTGTGTGTCATATA..... 602
71 eSerProAlaPheCysGlnIleTrpPheValArgGluAlaValArgThrTyrG 88

```

```

602 ..... 602
88 lyaatggluilemismethrleuapheulysleuaspheutylval 104
603 .....GCCCAT.....AATCG 613
105 leuprovalalanelleapgltyriletyrthrtrprthrlnaspar 121
614 AATGTGAGAAAGAACCCGTTCTTATGCGAACAAATCATGTCGGA 663
121 gmettrparglysahrtsertthrargalaglyserseuylserglyt 138
664 CAGACSTGATGACGACTTGTCTCAAAACSTGTCGAGGAGGACGA 713
138 hrspaleuasnargasnph...asplaelutprcyssettleglyala 153
714 TCCAGTTCTCATGTCSTGGAACSTGACSTGACSTTATGTCGAGTACA 763
154 Serasnasnprcysseuylthrtyrcysgllyseralalaglysergl 170
764 ACCGAAAGTAAAGCACTGCTGCTATGACGAAAGAAATATCAACAGA 813
170 ulysgluysersalvalalaspheulleargnsnleuaserseuyl 187
814 TTAAGCATCATGACATGATCATGATGATGATGATGATGATGATGAT 863
187 leuysalatytlethrthrleuileuileuileuileuileuileuile 203
864 TATTCSTATGACGAAAGTAAAGCAACGACGACGACGACGACGACG 913
204 tyrsertgluysersgluysleuproluysasnvalgluileuasnthr 220
914 ACCGAGTAAAGCACTGCTGCTATGACGAAAGTAAATATCAACGAT 963
220 ualalyglalyalvalalylysleu...alasertleuinsgltyrthr 236
964 ATACACATGACGACGACGACGACGACGACGACGACGACGACGACG 1013
236 tyrsertgluysersgluysalthrthrleuileuileuileuileu 252
1014 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1057
253 Aspsrtrpralatyarsnginglylleuysersertherrhneule 269
1058 .....AAC 1061
269 uatgaspalyglatgtyglrhevalleuproglyserglilleu 286
1062 CCACTGTAGAGAGCTTTGCGCGCTGCTGTAATAGCTGACGATGTC 1111
286 rothrtycsglulthrmetleualaleuyslytyvalthrserlyval 302
1112 ATTAGGAAGT 1123
303 leuglhileu 306
seq_name: SwissProt_40:CBPC_RAT
seq_documentation_block:
ID CBPC_RAT STANDARD: PRT; 309 AA.
AC P21961;
DT 01-AUG-1991 (Rel. 19, created)
DT 01-AUG-1991 (Rel. 19, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE A3).
GN CPA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE.

```

```

RX MEDLINE-91105153; PubMed-1988052;
RA Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA Neureath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules.";
RL Biochemistry 30:648-655(1991).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
DR PIR: A33118; A33118.
DR PIR: A38395; A38395.
DR HSSP: P09355; INSA.
DR MEROPS: M14.010; -.
DR InterPro: IPR000834; Zn_carpoept.
DR Pfam: PF00246; Zn_carpoept. 1.
DR PRINTS: PR00765; CROXYPRASA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL. 68 68
FT METAL. 71 71
FT METAL. 196 196
FT ACT_SITE 248 248
FT ACT_SITE 270 270
FT DISULFID 65 78
FT DISULFID 137 160
FT VARIANT 1 1
FT VARIANT 1 1
SO SEQUENCE 309 AA; 35786 MW; 20330FABC3EB83EF CRC64;

```

```

alignment_scores:
Quality: 567.50 Length: 303
Ratio: 2.925 Gaps: 6
Percent Similarity: 64.026 Percent Identity: 39.604

```

alignment_block:

US-09-980-881-1 x CBPC_RAT ..

Align seg 1/1 to: CBPC_RAT from: 1 to: 309

```

375 TATGACACGATGATGATGATGATGATGATGATGATGATGATGATGAT 424
||| ::|||::|||::|||::|||::|||::|||::|||::|||::|||::
8 tyralalystrylnasprtrpaspnlyllevalsertrprthrduysme 24
425 AACGTAGAGCATGCTGATGATGCTTACAAAATGCAATGATGATGAT 474
::|||::|||::|||::|||::|||::|||::|||::|||::|||::
24 tvalglulynhrprogluilevalserargyleysilleglyserthyr 41
475 TTGAGAGTACCCACCTGTATGTTTAAAGGTTTCTGGAAGAACAACA 524
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
41 algluaspasnproleuylvalleuylsile...glatglysasrgly 56
525 GCGAATAATGCCATATGATGATGATGATGATGATGATGATGATGAT 574
::|||::|||::|||::|||::|||::|||::|||::|||::|||::
57 gluatgylsalallepnehetaspysglylleuileuileuileuile 73
575 CTCTCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
::|||::|||::|||::|||::|||::|||::|||::|||::|||::
73 lserpralalphecysglintprphevaltyrglnalalalyserlytr 90
604 GCCATAT.....G 611
|| |||
90 llysasnasnillemetthrlyleuleuaspargmetasnphetyval 106
612 .....CG 613
107 leuprovalpheasnvalaspgltyriletyrthrtrprthrlyaspar 123
614 AATGTGAGAAAGAACCCGTTCTTATGCGAACAAATCATGTCGGA 663
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
123 gmettrparglysahrtsertlysasnprserseuylserglyt 140

```


RA Rees D.C., Lewis M., Lipscomb W.N.;
RT Refined crystal structure of carboxypeptidase A at 1.54-A
RL resolution.";
RN J. Mol. Biol. 168:367-387(1983).
[7]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX MEDLINE:96003618; PubMed:7556081;
RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT "The three-dimensional structure of the native ternary complex of
RT bovine pancreatic procarboxypeptidase A with propeptinase E and
RT chymotrypsinogen C.";
RL EMBO J. 14:4387-4394(1995).
RN
RP VARIANT ALLELIC.
RX MEDLINE:69283620; PubMed:5817619;
RA Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Identification of the amino acid replacements characterizing the
RT allotypic forms of bovine carboxypeptidase A.";
RL Biochemistry 8:2762-2768(1969).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC PROBROMELASE E.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -1- DATABASE: NAME-Worthington biochem.com/manual/C/COA.html".
CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: M61851; AAA30426.1; -;
DR EMBL: M61851; AAA30427.1; -;
DR EMBL: Z33906; CAB83955.1; -;
DR PIR: JN0126; CPBOA.
DR PIR: A31406; A31406.
DR PDB: 3CPA; 15-JAN-87.
DR PDB: 4CPA; 22-OCT-84.
DR PDB: 5CPA; 15-JAN-87.
DR PDB: 6CPA; 15-OCT-91.
DR PDB: 7CPA; 31-JAN-94.
DR PDB: 8CPA; 31-JAN-94.
DR PDB: 1CBX; 31-JAN-94.
DR PDB: 1CPS; 15-OCT-94.
DR PDB: 2CTB; 31-JAN-94.
DR PDB: 1ARL; 01-AUG-96.
DR PDB: 1ARM; 17-AUG-96.
DR PDB: 1BAV; 01-APR-97.
DR PDB: 1YME; 12-FEB-97.
DR PDB: 1CPX; 05-AUG-98.
DR PDB: 1PYT; 27-JAN-97.
DR MEROPS: M14.001; -;
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_caropept.
DR Pfam: PF00244; Propep_M14; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW 3D-structure; Polymorphism.
FT SIGNAL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A.
FT METAL 179 179 ZINC.
FT METAL 182 182 ZINC.

FT METAL 306 306 ZINC.
FT ACT_SITE 358 358 PROTON DONOR.
FT ACT_SITE 380 380 NUCLEOPHILE.
FT DISULFID 248 271
FT VARIANT 289 289
FT VARIANT 338 338
FT VARIANT 338 338
FT VARIANT 415 415
FT CONFLICT 95 95
FT CONFLICT 199 199
FT CONFLICT 203 203
FT CONFLICT 224 224
FT CONFLICT 232 232
FT CONFLICT 295 295
FT CONFLICT 295 295
FT TURN 114 114
FT TURN 119 120
FT TURN 125 138
FT HELIX 140 142
FT STRAND 143 150
FT STRAND 152 153
FT STRAND 156 162
FT STRAND 171 176
FT TURN 180 181
FT TURN 183 199
FT HELIX 200 202
FT TURN 204 212
FT STRAND 214 218
FT HELIX 223 231
FT TURN 232 232
FT TURN 234 235
FT STRAND 242 242
FT STRAND 244 245
FT STRAND 249 249
FT TURN 253 254
FT TURN 261 262
FT STRAND 267 267
FT TURN 270 271
FT TURN 273 274
FT STRAND 275 275
FT TURN 280 281
FT HELIX 284 296
FT STRAND 299 307
FT STRAND 311 314
FT TURN 324 325
FT HELIX 326 344
FT STRAND 349 352
FT HELIX 353 356
FT TURN 357 357
FT HELIX 364 370
FT TURN 371 372
FT STRAND 375 381
FT TURN 388 389
FT HELIX 393 416
SQ SEQUENCE 419 AA; 47082 MM; 21B86407B3BFC452 CRC64;

alignment_scores: Quality: 536.00 Length: 431
Ratio: 2.144 Gaps: 12
Percent Similarity: 58.005 Percent Identity: 32.715

alignment_block: US-09-980-881-1 x CBPA_BOVIN ..

Align seg 1/1 to: CBPA_BOVIN from: 1 to: 419

24 CTTTCAGCCTTCAGATGCTTACCAATGTTCTCTTGAGACAGCA 73
|||||
4 LeuLeuLeuLeuSerValLeuLeuGlyAlaAlaLeuGlyLysGluAsp.. 19
74 TGTCTTCGCTTCAGAGTGGCCAGTTCAGCTGCTTCTAGAACCT 123
|||||
20PheValGlyHisGlnValLeuArgIleThrAlaAlaAspG 33

```

124 CTAGCAAGCTTCAGTCTTACAGAACTTATACAAACATATGAGATTGTT 173
   :: ::::::::::::::::::::
33 LAAAGLVALGlnThrValLysGlnLeuGlnAspLeuGlnHisLeuGln 49
174 CTC.....TGG.....CAGCGGTAACAGCTGACCTTATGCT 205
   ||| ||| |||
50 LeuAspPheTrpArgGlyProGlyGlnPro..... 59
206 GAAGAAACAAAGATCCATTTTGTGTAATGATCATCTGATGTCGACAAAG 255
   :: ::::::::::
60 ... GlySerProIleAspValArgValProPheProSerLeuGlnAlaVal 75
256 TGAAGACCCATTTAAATGTGACGCGCAATTCATGACAGTGTCTTGCGCA 305
   ||| ::::::::::
75 AluValAlaPheLeuGlnAlaHisGlyIleArgTyrArgIleMetIleGln 91
306 GACGTGGGAAGATCTTAT.....CAACAGACATTTCCACAGACAC 346
   ||| ::::::::::
92 AspValGlnSerLeuLeuAspGlnGlnGlnGlnGlnMetPheAlaSerGln 108
347 AGTCAGCCCCGAGCTCCGATCGTAC...TATGAACAGATATCACTAC 393
   ||| ::::::::::
108 nSerArgAlaArgSerThrAsnThrPheAsnTyrAlaThrTyrHisThrL 125
125 euAspGlnIleTyrAspPheMetAspLeuValAlaGlnHisProGln 141
444 ATGCTTACAAATAATCCATGATGCTCATTTGAGAGATGCCACTGTA 493
   ::::::::::::::::::::
142 LeuValSerLysLeuGlnIleGlyArgSerTyrGlnGlyArgProIleTyr 158
494 TGTTTTAAAGTTTCTGAAAAAGAACACAGCAAAATGTCATATGGA 543
   ||| ::::::::::
158 rAlaIleuLysPheSer...ThrGlySerAsnArgProAlaIleTrpL 174
544 TTGACTGTGGATCCATGCCACAGAAATGATCTCTCTGCTTCTGCTTG 593
   ||| ::::::::::
174 LeuAspLeuGlyIleHisSerArgGlnTrpIleThrGlnAlaThrGlyVal 190
594 TGGTTCATA..... 602
191 TrpPheAlaLysLysPheThrGlnAspTyrGlnGlnAspProSerPheThr 207
602 ..... 602
207 rAlaIleLeuAspSerMetAspIlePheLeuGlnIleValThrAsnPro 224
603 .....GCCCATATTCGAATGTGGAGAAAGAACCGT 632
224 spGlyPheAlaPheThrHisSerGlnAsnArgLeuTrpArgLysTrpArg 240
633 TCTTTCATGCGAAACATGATCATCTGCAAGACAGCTGAAATGCAACT 682
   ||| ::::::::::
241 SerValThrSerSerSerLeuGlyValAlaAspAlaAsnArgAsnTr 257
683 TGTCTCCAAACACTGGTGTGAGAGAGATGATCACTCTCATGCTCGG 732
   :: ::::::::::
257 P...AspAlaGlnPheGlyLysAlaGlnAlaSerSerProCysSerG 773
773 AAACCTACCTGAGCTTTATCTGACTCAGAACACAGAGAGAGAGAGAG 782
   ||| ::::::::::
273 LThrTyrHisGlyLysTyrAlaAsnSerGlnValGlnValLysSerIle 289
783 GCGAGTGTCTTGAGAAATATCAACAGATTAAGCATATCATGAGAT 832
   ::::::::::::::::::::
290 ValAspPheValLysAspHisGlyAsn...PheLysAlaPheLeuSerIle 305
833 GCATTCATATCCCGACATATAGTGTTCATATTCCTATACAGAAAGTA 882
   ::::::::::::::::::::
305 eHisSerTyrSerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrGln 322
883 AAAGCAAAAGACCATGAGAACTCTCTACTAGCCACTGAAAGAGTTGCT 932

```

```

:: ||| ||| :::::::::: ||| :::::::::: ||| ::::::::::
322 eXlleProAspLysTrpThrGlnLeuAsnGlnValAlaLysSerAlaValGln 338
933 GCTATTGACAAAACCTAGTAAATAATACAGATATACATGCGCATGCGTC 982
   ||| :::::::::: ||| ||| |||
339 AlaLeu...LysSerLeuTyrGlyThrSerTyrLysTyrGlySerIleI 354
983 AGAAACCTTATACCTGATCTCTGAGAGTGGGAGAGATGGATCATATGAT 1032
   ||| :::::::::: ||| ||| ::::::::::
354 eThrThrIleTyrGlnAlaSerGlyLysSerIleAspTrpSerTyrAsnG 371
1033 TGGGCAATCAAAATTTGCTTACATCAA..... 1060
   ||| :::::::::: ||| |||
371 IIndIleLysTyrSerPheThrPheGlnLeuArgAspThrGlyArgTyr 387
1061 .....CCGACCTGAGAGAGACTTT 1080
   ||| :::::::::: ||| |||
388 GlyPheLeuLeuProAlaSerGlnIleIleProThrAlaGlnIndThrTr 404
1081 TGGCGCTGTCTTAAATAAGCTTGCAATGATCATATGAGATGTT 1123
   :::::::::: ||| ||| ::::::::::
404 PleuGlyValIleuThrIleMetGlnHisThrLeuAsnAsnLeu 418
seq_name: SwissProt_40:CBPL_RAT
seq_documentation_block:
ID CBPL_RAT STANDARD: PRT; 419 AA.
AC P00731;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cardoxyzepidase A1 precursor (EC 3.4.17.1).
GN CPAL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82105986; PubMed=6275388;
RA Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standing D.N.,
RA Picket R.L., Valenzuela P., Rutter W.J.;
RT "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
RT characterization of the gene."
RT Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase gene
RT family."
RL J. Biol. Chem. 263:17837-17845(1988).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01232; CAA24542.1; -
DR EMBL; J00713; AAA40893.1; -
DR EMBL; M23990; AAA40955.1; -
DR EMBL; M23960; AAA40955.1; JOINED.
DR EMBL; M23985; AAA40955.1; JOINED.
DR EMBL; M23986; AAA40955.1; JOINED.

```



```

684 GTCTCCAAACACCTGCTGAGAGGATCATGCTTCATCTCGCA 733
    ::::: |||||::: |||||:::
256 ..AspIaagIphneGlygIyProGlyAlaSerSerAsnProCysSerAs 271
734 AACCTAGCTAGCTTATCTGATGATGATGATGATGATGATGATGATG 783
    ::::: |||||::: |||||:::
271 pSerTYHISGILProSerAlaAsnSerGluValGluValIleSerIleV 288
784 CTAGTCTTGTGAGAGAAATATCAACGATTAAGCATATCATCATGATG 833
    ::::: |||||::: |||||:::
288 alaPheIleIleSerHis...GlyValIleValIlePheIleIleLeu 303
834 CATTCATACCTCCAGCATATAGTCTTCCATATTCCTATACACAGTA 883
    |||||::: |||||::: |||||:::
304 HisSerTYSerGlnLeuIleMetPheProTYGlyTYIleCysTYHly 320
884 AAGCAAGACCATGAGCACTGCTCTAGTACGAGGATGATGATGATG 933
    |||||::: |||||::: |||||:::
320 sIeuAspAspPheAspGlnLeuSerGluValAlaGlnIleValIleGln 337
934 CTATGTGCAAACTAGTAAATATCAAGTATACATGATGATGATGATG 983
    ::::: |||||::: |||||:::
337 erIeu...ArgSerLeuHISGlyThrIleTYIleValGlyProIleCys 352
984 GAAACCTTATACCTAGCTCTGAGGAGGAGGATGATGATGATGATG 1033
    ::::: |||||::: |||||:::
353 SerValIleTYGlnAlaSerGlyIleSerIleAspIleSerTYIleAsp 369
1034 CGGATCAAAATATCTGTTTACATCAAAAC..... 1061
    |||||::: |||||::: |||||:::
369 rGlyIleIleTYIleSerPheAlaPheGlnLeuAlaGspIleTYIleG 386
1062 .....CCACCTGTAGAGAGCTTTGCGC 1085
    |||||::: |||||::: |||||:::
386 IyPheLeuLeuProAlaIleGlnIleLeuPro 396
seq name: SwissProt_40:CBPL_HUMAN

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X67318; CAA47732.1; -
DR PIR: S02810; S02810.
DR PIR: S29127; S29127.
DR HSSP: P00730; 1PPT.
DR MEROPS: M14.001; -.
DR MTM: I14850; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_Cardopept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_Cardopept; 1.
DR PRINTS: PR00765; CARBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
DR Hydrolase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT STGNPL 1 16
FT PROPEP 17 110
FT CHAIN 111 419
FT METAL 179 179
FT METAL 182 182
FT METAL 306 306
FT ACT_SITE 358 358
FT ACT_SITE 380 380
FT DISULFID 248 271
SQ SEQUENCE 419 AA; 47140 MW; 439FAFFAE958B1 CRC64;

alignment_scores:
    Quality: 489.50      Length: 401
    Ratio: 2.128        Gaps: 13
    Percent Similarity: 57.357    Percent Identity: 32.668

alignment_block:
US-09-980-881-1 x CBPL_HUMAN ..

Align seg 1/1 to: CBPL_HUMAN from: 1 to: 419

24 CTTTCAGACCTTGACCTTGTACCAATGTTCTCTCTGAGCAGCA 73
    ||| |||||::: |||||::: |||||:::
4 LeuLeuValIleSerValLeuLeuGlnValAlaValPheGlyIleGlnAsp.. 19
74 TGTCTTCCGCTTCAGAGTGCCCAAGTTCTAGCTGCTTCTCTAGAACCT 123
    ||| ::||| |||||:::
20 .....PheValGlyHISGlnValIleuArgIleSerValAlaAspG 33
124 CTAGCAGATTCAGTCTCTCAGAAATCTTACTACACATATGACATGTT 173
    ::||| |||||::: |||||::: |||||:::
33 IuAlaGlnValGlnIleValIleValIleValIleValIleValIleVal 49
174 CTC.....TGACAG.....CCGATGACAGCTGA 196
    ||| |||||::: |||||::: |||||:::
50 LeuAspPheIlePArgGlyProAlaHisProGlySerProIleAspVal.. 65
197 CTTATTGTGAAGAAAAACAAGTCAATTTTGTAAATGCAATGCTGATG 246
    ::||| |||
66 .....ArgValProPhe.....ProSerI 72
247 TCGCAATGTGAAGCCATTTAATGTGAGGGAATTCATGACGTGTC 296
    ::||| |||||::: |||||::: |||||:::
72 IeGlnAlaValIleValIlePheLeuGlnSerHisGlyIleSerTYGlyThr 88
297 TTGCTGCAGACAGCTGAGATCTTATT.....CAACAGCAGATT.. 335
    ::||| |||||::: |||||::: |||||:::
89 MetIleGlnAspValGlnSerLeuLeuAspGlnGlnGlnGlnMetPhe 105
36 .....TCCACGACACAGTCAAGCCCGAGGCT 363

```



```

54 rpylserProSerPheAsnArgProValAspValIleValProSer 70
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
213 AAACAATCCATTTTGTAAATGCATGATGTCGACAATGGAAGC 262
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
71 ValSerLeuAlaPhe.....LysSe 78
263 CCATTAAATGTAGCGGAATTCACAGTGTCTGTCGACAGCTGG 312
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
78 rPheLeuArgSerGlnGlyLeuGluValValThrIleGluAspLeuG 95
313 AAGATCTATT.....CAACAGAGATTTCACAGCAGACGTCAGC 353
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
95 InAlaLeuLeuAspAsnGluAspSerGluMetGlnHisAsnGluGln 111
354 CCCCAGACCTCCGACATCGTAC...TATGACAGATGATCAGTAAATGA 400
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
112 GluArgSerSerAsnAsnPheAsnGlyAlaValThrHisSerLeuGluAl 128
401 AATCTATTCTTGATGATTAATTAATCTAGAGAGGATCCTGATATGCTTA 450
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
128 aileTyHisGluMetAspAsnIleAlaAlaAspPheProAspLeuAla 145
451 CAAAATCCSACATTGATCTCTCATTTGAGAAAGTACCCACCTATGTTTA 500
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
145 rGaArgValIleGlyHisSerPheGluAsnArgProMetGlyValLeu 161
501 AAGCTTCTGGAAGAAAGCAACAGCCAAATATGCATATGATGATGCTG 550
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
162 LysPheSerThrGlyValArgValArgProAlaValThrLeuAlaAl 178
551 TGGATTCATGCCAGAGATGATCTCTGCTTCTGCTGTTGG... 596
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::
178 aGlyIleHisSerArgIleuTrpIleSerGlnAlaThrAlaIleTrpPhr 195
596 ..... 596
195 laArgIleValSerAspTyrGlnArgAspProAlaIleThrSerIle 211
597 .....TTCATA..... 602
212 LeuGluMetAspIlePheLeuProValAlaAsnProAspGlyTy 228
603 .....GCCATTAATGATGTGAGAAAGAACCGTCTTCTTCT 639
228 rValIleTyrGlnThrGlnAsnArgLeuTrpArgIleThrArgSerArg 245
640 ATCGGAACATCATGTCATGCGAAGACCTGATATGACATTTGCTGCC 689
245 snProGlySerSerCysIleGlyAlaAspProAsnArgAsnTrp...Asn 260
690 AAACACTGCTGTAGAGAGGTGCATGCTCCTCATGCTCGAAGACCTA 739
261 AlaSerPheAlaGlyLysGlyAlaSerAspAsnProCysSerGluValTy 277
740 CTGTGACTTATCTGATGACAGACCAAGAGGAGGAGGCTGCTAGTT 789
277 rHisGlyProHisAlaAsnSerGluValGluValLysSerValValAsp 294
790 TCTTGAGAAATATCATCAGATTAAACATATACATACACATGATCATCA 839
294 heIleGlnIleHisGlyAsn...PheLysGlyPheIleAspLeuHisSer 309
840 TACTCCAGCATATAGTGTTCATATTCCTATATACAGACGATAAAGCAA 889
310 TyrSerGlnLeuLeuMetIleTyrProTyrGlyTyrSerValLysAlaI 326
890 AGACCATGAGGAAGTGTCTAGTACGACATGAGAGAGTTCGTCATTTG 939
326 asPAlaGlnGluLeuAspLysValAlaArgLeuAlaAlaLysAlaLeu 343
940 ACAAACTAGTAAATATACAGATATACATGCGCATGGCTCGAAGACC 989
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::

```

```

343 laSerValSer...GlyThrGluTyrGlnValGlyProThrCysThrThr 358
990 TTATACCTAGCTCCTGAGGTGGAGACGATGATGATGATTTGGGCAAT 1039
359 ValTyrProAlaSerGlySerSerIleAspTrpAlaTyrAspAsnGlyI 375
1040 CAAATATTCGTTTACA 1055
375 elysPheAlaPheThr 380
seq_name: SwissProt_40:CBPZ_SIMV1

```

```

seq_documentation_block:
ID CBPZ_SIMV1 STANDARD; PRT; 304 AA.
AC P42788;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.**) (Fragment).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cut;
RX MEDLINE=94093864; PubMed=8269093;
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins."
RL Insect Mol. Biol. 1:149-163(1993).
CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08481; AAL18531.1; -.
DR HSSP: P48052; IAYE.
DR InterPro: IPR000834; Zn_carboxpept.
DR Pfam: PF00246; Zn_carboxpept. 1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 58 58 ZINC (BY SIMILARITY).
FT METAL 61 61 ZINC (BY SIMILARITY).
FT ACT_SITE 184 184 ZINC (BY SIMILARITY).
FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 125 148 BY SIMILARITY.
SQ SEQUENCE 304 AA; 34849 MW; 25E53FF8A6A9144 CRC64;

```

```

alignment_scores:
Quality: 409.00 Length: 308
Ratio: 2.260 Gaps: 10
Percent Similarity: 58.766 Percent Identity: 31.818

```

alignment_block:

US-09-980-881-1 x CBPZ_SIMV1 ..

Align seg 1/1 to: CBPZ_SIMV1 from: 1 to: 304

381 CAGTATCACTCATTAATGAATCTATTCTTGATAGATTATTAAGTGA 430

OM of: US-09-980-881-1 to: SPTREMBL_19:* out_format : pfs
Date: Sep 18, 2002 4:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+22p_model -DEV=x1h
-O=/cgn2.1/USPTO_sptool/US09980881/runatc_16092002_140041_8910/app-query.fasta_1.1655
-DB=SPTREMBL_19 -OEMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPTOL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -TGAPOP=10.000 -TGAPEXT=0.500
-DELDP=6.000 -DELEXF=7.000 -START=1 -MATRIX=blsnum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881.ecgnl.1.101 -NCPU=6 -ICPU=3 -LONGLOC
-DEVTIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-1
Query length: 1573
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 17294929
Search time (sec): 93.320000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Document
sp_human:015114	1885.00	3342.23	1.9e-178	423	015114 homo sapiens (human) . pc
sp_human:092PY6	1884.00	3342.01	2.3e-178	360	092PY6 homo sapiens (human) . ca
sp_human:0961Y4	1880.00	3333.31	6.0e-178	423	0961Y4 homo sapiens (human) . ca
sp_rodent:093JH6	1592.50	2820.48	2.2e-149	422	093JH6 mus musculus (mouse) . ca
sp_rodent:09QZFO	1586.50	2809.17	8.8e-149	422	09QZFO mus musculus (mouse) . ca
sp_rodent:095QV9	1568.50	2777.66	5.4e-147	422	095QV9 rattus norvegicus (rat) . ca
sp_human:09NT18	1029.00	1822.61	1.8e-93	198	09NT18 homo sapiens (human) . da
sp_vertebrate:09PUF2	714.00	1253.50	4.3e-62	416	09PUF2 boehrs jaraeca (jaraec)
sp_human:096B94	652.50	1143.77	5.6e-56	417	096B94 homo sapiens (human) . st
sp_human:096B08	646.00	1132.20	2.5e-55	416	096B08 homo sapiens (human) . hy
sp_mammal:09ASBP3	642.50	1126.03	5.3e-55	419	09ASBP3 sus scrofa (pig) . proca
sp_rodent:0975Y7	531.00	930.93	5.9e-44	279	0975Y7 rattus norvegicus (rat) . ca
sp_rodent:09CVD1	516.50	901.12	1.8e-42	419	09CVD1 mus musculus (mouse) . 22
sp_mammal:09TV85	511.00	891.35	6.4e-42	417	09TV85 sus scrofa (pig) . carbox
sp_human:096ON3	489.50	852.95	6.4e-42	417	096ON3 homo sapiens (human) . ca
sp_human:09B612	489.50	852.95	6.4e-42	417	09B612 homo sapiens (human) . hy
sp_human:09B617	489.50	852.95	6.4e-42	419	09B617 homo sapiens (human) . ca
sp_human:09B629	484.00	848.27	2.7e-39	247	09B629 homo sapiens (human) . ca
sp_invertebrate:09VL87	460.00	800.21	7.5e-37	424	09VL87 drosophila melanogaster . hy
sp_invertebrate:09W475	459.00	788.40	1.2e-36	1192	09W475 drosophila melanogaster . hy
sp_invertebrate:09VL22	451.50	785.10	5.2e-36	422	09VL22 drosophila melanogaster . hy
sp_invertebrate:09VTH1	436.00	754.30	2.0e-34	584	09VTH1 caenorhabditis elegans
sp_invertebrate:09W478	418.00	724.93	1.1e-32	440	09W478 drosophila melanogaster . hy
sp_invertebrate:09V421	407.50	706.68	1.2e-31	419	09V421 drosophila melanogaster . hy
sp_invertebrate:09V566	402.00	696.75	4.3e-31	424	09V566 drosophila melanogaster . hy
sp_invertebrate:09V186	402.00	696.61	4.3e-31	430	09V186 drosophila melanogaster . hy
sp_invertebrate:09VTH2	400.50	691.73	6.5e-31	540	09VTH2 caenorhabditis elegans
sp_invertebrate:09V9K2	399.50	692.22	7.7e-31	427	09V9K2 aedes aegypti (yellow)
sp_invertebrate:019121	397.00	690.47	1.3e-30	323	019121 caenorhabditis elegans
sp_invertebrate:09V432	396.00	685.41	1.7e-30	453	09V432 drosophila melanogaster . hy
sp_invertebrate:061532	386.50	669.11	1.5e-29	423	061532 drosophila melanogaster . hy
sp_invertebrate:023318	382.50	658.91	4.0e-29	581	023318 caenorhabditis elegans
sp_invertebrate:09VX86	370.00	628.77	7.9e-28	1430	09VX86 drosophila melanogaster . hy
sp_invertebrate:09N3S6	370.00	627.23	7.9e-28	545	09N3S6 caenorhabditis elegans
sp_invertebrate:09VCM8	368.50	636.05	9.4e-28	467	09VCM8 drosophila melanogaster . hy
sp_invertebrate:09VTC6	362.00	623.60	3.7e-27	455	09VTC6 caenorhabditis elegans
sp_invertebrate:09XU75	358.00	613.88	1.1e-26	666	09XU75 caenorhabditis elegans
sp_invertebrate:09V564	355.00	613.12	2.0e-26	415	09V564 drosophila melanogaster . hy
sp_invertebrate:09V567	354.50	612.58	2.2e-26	400	09V567 drosophila melanogaster . hy
sp_invertebrate:09V568	344.50	593.93	2.1e-25	354	09V568 drosophila melanogaster . hy

sp_invertebrate:0961J8 + 344.50 594.32 2.2e-25 418 | 0961J8 drosophila melanogaster . hy
sp_invertebrate:09VR3 + 337.00 583.77 1.2e-24 312 | 09VR3 drosophila melanogaster . hy
sp_human:09NR19 + 334.00 578.33 2.3e-24 315 | 09NR19 homo sapiens (human) . ca
sp_vertebrate:09QW12 + 324.50 567.89 1.7e-23 161 | 09QW12 brachydanio rerio (zebrafish)
sp_fungi:014418 + 320.00 550.61 6.1e-23 418 | 014418 metarhizium anisopliae . hy

seq_name: sp_human:015114

seq_documentation_block:

ID: 015114 PRELIMINARY; PRT; 423 AA.
AC: 015114;
DT: 01-NOV-1996 (TREMBL:rel. 01, Created)
DT: 01-NOV-1996 (TREMBL:rel. 01, Last sequence update)
DE: 01-DEC-2001 (TREMBL:rel. 19, Last annotation update)
DE: PCPB PROTEIN.
GN: PCPB.
OS: Homo sapiens (Human).
OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC: Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX: NCBI_TaxID=9606;
RN: [1]
RP: SEQUENCE FROM N.A.
RC: TISSUE=LIVER.
RX: MEDLINE-92042093; PubMed-1939207;
RA: Eaton D.U., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT: "Isolation, molecular cloning, and partial characterization of a novel
RT: carboxypeptidase B from human plasma."
RL: J. Biol. Chem. 266:21833-21838(1991).
DR: EMBL: M75106; AAA60042.1; .
DR: HSSP: P00730; 1PVT.
DR: MEROPS; M14.009; .
DR: InterPro; IPR003146; Propep_M14.
DR: Pfam; PF02244; Propep_M14; 1.
DR: Pfam; PF00246; Zn_carboxypep. 1.
DR: PRINTS; PR00765; CRBOXYPRASEA.
FT: CHAIN: 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ: SQUENCE 423 AA; 48442 MW; 9B383272F6E79E4 CRC64;

alignment_scores:

Quality: 1885.00 Length: 423
Ratio: 5.122 Gaps: 2
Percent Similarity: 86.998 Percent Identity: 86.525

alignment_block:

US-09-980-881-1 x 015114

Align seg 1/1 to: 015114 from: 1 to: 423

18 ATGAAGCTTTGAGCGCTTGCAGTCCCTTGACCCATTGCTCTCTGCA 67
|||||
1 MetylsleucysSerLeuAlaValLeuValProIleValLeuPhcysG1 17
68 GCAGCATGCTTTCGCGCTTCACAGTGGCCCAAGTTAGCTGCTTCCGA 117
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaValLeuProA 34
118 GAACCTTACGCAAGTCAAGTTCATACGAATCTTACATACCAATATAG 167
|||||
34 rGThnSerArgGlnValGlnValLeuGlnAsnLeuThrThrTrpGlu 50
168 ATTGTTCTCTGCGACCGGTAAACAGCTGACCTTATGTAAGAAAAACA 217
|||||
51 IleValLeuIleuTrpGlnProValThrAlaPheLeuIleValIleGlyG1 67
218 AGTCCATTTTGTGTAATGCAATCTGATGTCGACAAATGTAAGCCCAT 267
|||||
67 nValHisPhePheValAsnAlaSerAspValAlaPheValIleValHisL 84
268 TAAATGTAGCGGAATTCATGCAATGCTGCTGCTGCGACAGTGAAGAT 317
|||||
84 euAsnValSerIleProCysSerValLeuLeuAlaPheValIleValL 100

```

318 CTTATTCAACACAGATTTCCACACAGCTGACCCCCGAGCCTCCGC 367
|||||
101 Ieulieglnnglnlleserbsnspthrlvalserproarglaserl 117
|||||
368 ATCTACTATGAAACAGTACTACTAAATGAATCTATCTTGAGAG 417
|||||
117 aserlytrgluglntrhlsrserleusnseulterlysertrileg 134
|||||
418 AATTTATAGTAGAGGACATCCTGATATGCTTACAAAAATCCACATGGA 467
|||||
134 IupheilethrcIuarghlsproaspmetleuthrlsilestleegly 150
|||||
468 TCCCATTTGAGAAAGTACCCACTATATGTTTAAAGTTTGGAAAAA 517
|||||
151 SerSerPheglulysrtrproleuylvalleuylvalserglylys 167
|||||
518 ACAAAACACAAAAATGCCATATGATGATGATGATGATGATGATGATG 567
|||||
167 uclnhrhlaIysasnalailetrpilaaspCysglylehlslalarg 184
|||||
568 AATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 608
|||||
184 Iutrpileserproalaphcysleutrpheileglyhlsllethrgln 200
|||||
608 ..... 608
201 PheYrGlyleileglylnlYrthrasnleuArlgleuValasprh 217
|||||
608 ..... 608
217 eYrValmetProvalValasnValasprglyYrthrasprYsertrpYsl 234
|||||
609 ..AATCGAATGTGGAGAAAGACCGTCTTTCTATGCGAACAATCATGTC 656
|||||
234 ysaasArGmetrParglysaasArGserPheYrthrasnlnhscys 250
|||||
657 ATCGAAGACAGCTGAATAGCACTTTGTCSCAAACATCGTGTGAGGA 706
|||||
251 IleelythraspleuasnArGasnPhehlaserYshstrCysglul 267
|||||
707 AGGTGATCCAGTTCTCATGCTCGGAACCTACTGAGATTATCTCTG 756
|||||
267 uGlyAlaserSerSerSerSerSerGluThrYrCysglYleuYrtrpG 284
|||||
757 AGTAGAACAAGAGTAGAGGACAGTGTCTTGTGAGAAATATC 806
|||||
284 IuSerGluProGluValIysAlaValAlaserPheleuArGargAsnIle 300
|||||
807 AACGAGATTAAAGCATATCAGCATGATTCATCTCCGAGCATATAGT 856
|||||
301 AsnGlnIleYsalatYrIleSerMetIleSerYserGlnIleIleVa 317
|||||
857 GTTTCATATTTCTATACAGCAAGTAAAGCAAGACCATGAGAACTGT 906
|||||
317 IPherProYrSerYrthrasrYserYsAsprhISgluldeus 334
|||||
907 CTCTAGAGGACAGTAGAGGACAGTGTCTATGACAAATAATGTAATAAT 956
|||||
334 erleuValAlaserGluAlaValAlaIleGluYstrSerYsAsn 350
|||||
957 ACCAGATATACATGCGCATGGCTCAGAAACCTTATACCTAGTCTCTG 1006
|||||
351 ThrArgYrthrhlsISglYserGluThrleuYrleuAlarProG 367
|||||
1007 AGGTGGGAGCATGATGATGATGATGATGATGATGATGATGATGATG 1051
|||||
367 YglYglYAspArPrlleYrthraspleuGlyleYstrSerPheThrI 384
|||||
1052 .....TAC 1054
384 IecluleuArGAsprhnglyYrthYrGlyrPheleuLeuProGluArGTYr 400
|||||

```

```

1055 ATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGCTCTCTAAATAGCTTG 1104
|||||
401 IleYsProthrCysArgGluAlaPhehlalalValserYsIleAlatr 417
|||||
1105 GCARGTCTATTGGAATGTT 1123
|||||
417 phlValleleArGAsnVal 423
|||||
seq_name: sp_human:Q9P2Y6

```

```

seq_documentation_block:
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B-LIKE PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBL_TaxID-9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92042093; PubMed-1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma."
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL: AB011969; BAA90475.1;
DR HSSP: P00730; 2CTC.
DR MEROPS: M14.009;
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept; 2.
DR PRINTS: PR00765; CRHOXYPTASEA.
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

```

```

alignment_scores:
Quality: 1884.00 Length: 360
Ratio: 5.263 Gaps: 0
Percent Similarity: 99.444 Percent Identity: 98.333

```

```

alignment_block:
US-09-980-881-1 x Q9P2Y6 ..

```

```

Align seg 1/1 to: Q9P2Y6 from: 1 to: 360

```

```

18 ATAGACGTTTGACGCTTGACGCTTTGTACCCATTTGTTCTCTGTGA 67
|||||
1 MetIlyleuYsSerleuAlaValleuValProIleValleuPheYsgl 17
|||||
68 GCAGCATGCTTCCGCTCCAGAGTGCGCAAGTTCTAGCTGCTCTCTTA 117
|||||
17 uGlnIhlsvalPheAlaPheGlnthrcIynlvalleuAlaIalaleuProA 34
|||||
118 GAACCTCTAGGCAAGTTCAAGTTCTACAGAACTTACTACAAATATGAG 167
|||||
34 rghrSerArGlnValGlnValleuGlnAsnleuThrhrthYrGlu 50
|||||
168 ATGTCTCTCTGGACCGGTAACAGCTGACCTTATTTGTAAGAAAAACA 217
|||||
51 IlevallleuTrpelnProvalThralaAspleuIlevalIlylsYsgl 67
|||||
218 AGTCATTTTGTAAATGATGATGATGATGATGATGATGATGATGATGAT 267
|||||

```

```

67 nvalhisphphevalasnalaserpvalaspvallysalaahisl 84
268 TAAATGTGACGGGAATTCATGACAGTGTCTGTGTCGACAGCTGGAGAT 317
      |||
84 euaasvalserglylleprocyservalleuleualaaapvalglasp 100
      |||
318 CTATTTAACAAGCAGATTTCACACAGACAGTCAAGCCCGGAGCTCCGC 367
      |||
101 leuileglinginglinleserasnpsprhvalserproaaglaaseral 117
      |||
368 ATGCTACTATGACAGATATCATCTCAATTAATGATCTATCTGGATAG 417
      |||
117 asertytyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 134
      |||
418 AATTTATATGAGAGCAGATCTGATATGCTTACAAAATCCACATTGGA 467
      |||
134 lupheillethngluarhishisproaspmetleuthrlysllehislllegly 150
      |||
468 TCCTCATTTGAGAAAGTACCCACCTGATGTTTAAAGTTTCTGAAAAGA 517
      |||
151 SerSerPheglulysyltyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 167
      |||
518 ACAAAACAGCCAAAATGACATATGATGATGATGATGATGATGATGATG 567
      |||
167 uclnalaialalsasnaiailetrpilleaspcysglyllehislaaarg 184
      |||
568 AATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 617
      |||
184 lutrpiieserproialaphecyseuotrphelieglyhisasnargmet 200
      |||
618 TCGAAGAAAGAACCGTTCTTTCTATGCAACATCATTTGATCGAAGACA 667
      |||
201 TTPATglysasnargserphetyrfaalsasnhiacyllleglythas 217
      |||
668 CCTGATATGACACTTGTCTCCAAAACCTGCTGTGAGGAAGAGTGACATCA 717
      |||
217 pheuasnarqsnphhealaserlyshistpysgluglulysalaser 234
      |||
718 GTTCCTCATGCTCGAAACCTACTGAGACTTATCTCTGAGTCAAGACA 767
      |||
234 ersersecyssergluthrtyrcysglyleutyrrprogluserglupro 250
      |||
768 GAACTGAGCAGAGTGGCTAGTTTCTTGAGAAATAATCAACAGATTAA 817
      |||
251 GlulvalysalaValalaserPheleuargragnliasnlinlely 267
      |||
818 AGCTATACATGACATGATCATCTCCAGCATATAGTGTTCATATT 867
      |||
267 salatyrilleermethisserlyrserglinhisillevalrphero 284
      |||
868 CTTATACAGCAAGTAAAGCAAGACCATGAGAACTGTCTAGTAGCC 917
      |||
284 ertyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 300
      |||
918 AGTGAACAGATGCTGCTATTTGACAAAATAGTAAATACAGATATAC 967
      |||
301 SerGlulvalalargalaileglulysilleserlysnthrargtyrth 317
      |||
968 ACATGGCAGTGGCTCAGAAACCTTATACCTGATCTCGAGAGTGGAGCG 1017
      |||
317 rhisgllyhisgllysergluthrtyrtyrtyrtyrtyrtyrtyrtyr 334
      |||
1018 ATTGATCTATGATTTGGGACATCAATATTTGTTATCATCAACCCACT 1067
      |||
334 sprtpletyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 350
      |||
1068 GTAGAGAGCTTTTGGCGCTCTCTATAA 1097
      |||
351 ValGlulysleuileuproleuSerleu 360
seq_name: sp_human:Q961Y4
seq_documentation_block:

```

```

ID 0961Y4 PRELIMINARY; PRT; 423 AA.
AC 0961Y4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE B2 (PLASMA).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKETELAL MUSCLE;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC007057; AAH07057.1; -.
SQ SEQUENCE 423 AA: 48412 MW: 9B383E03F6E7CF5 CRC64;

```

```

alignment_scores:
  Quality: 1880.00 Length: 423
  Ratio: 5.109 Gaps: 2
Percent Similarity: 86.998 Percent Identity: 86.288

```

```
alignment_block:
us-09-980-881-1 x 0961Y4 ..

```

```
Align seg 1/1 to: 0961Y4 from: 1 to: 423

```

```

18 ATGAAGCTTTCGACGCTTGACGCTCTTGACCATTTGTTCTTCTGTGCA 67
      |||
1 Metlysleucysserleualavalleualproillevalleuphecygl 17
      |||
68 GCAGCATGCTTGGGCTTCAGAGTGGCCAGTGTACTGCTGCTCTCTA 117
      |||
17 uclnhisvalphealaphheglinseryglvalleualalaaleuproa 34
      |||
118 GAACCTCTAGGCAAGTTCAAGTTCTACAGAACTTACTACACATATGAG 167
      |||
34 rglthserarglnvalglvalleuglnaasnleuthrthtyrtyrtyr 50
      |||
168 ATTGTTCTCTGCGACCGCTTAACAGCTGACCTTATTTGAAAGAAAACA 217
      |||
51 llevalleuthrpglnprovalthralaspleuilevallysllysyl 67
      |||
218 AGTCCATTTTGTGTAATGCATCTGATGTCGACAAATGGAAGCCCAT 267
      |||
67 nvalhisphphevalasnalaserpvalaspvallysalaahisl 84
      |||
268 TAAATGTGACGGGAATTCATGACAGTGTCTGTGTCGACAGCTGGAGAT 317
      |||
84 euaasvalserglylleprocyservalleuleualaaapvalglasp 100
      |||
318 CTATTTAACAAGCAGATTTCACACAGACAGTCAAGCCCGGAGCTCCGC 367
      |||
101 leuileglinginglinleserasnpsprhvalserproaaglaaseral 117
      |||
368 ATGCTACTATGACAGATATCATCTCAATTAATGATCTATCTGGATAG 417
      |||
117 asertytyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 134
      |||
418 AATTTATATGAGAGCAGATCTGATATGCTTACAAAATCCACATTGGA 467
      |||
134 lupheillethngluarhishisproaspmetleuthrlysllehislllegly 150
      |||
468 TCCTCATTTGAGAAAGTACCCACCTGATGTTTAAAGTTTCTGAAAAGA 517
      |||
151 SerSerPheglulysyltyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 167
      |||
518 ACAAAACAGCCAAAATGACATATGATGATGATGATGATGATGATGATG 567
      |||
167 uclnalaialalsasnaiailetrpilleaspcysglyllehislaaarg 184
      |||

```

```

568 AATGATCTCTCTGCTTCTGCTTGTGTTATAGGCAT..... 608
184 IuPrIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
608 ..... 608
201 PheTyLeIleIleGlyGlnTyTrhAsnLeuLeuArgLeuValAspH 217
608 ..... 608
217 eTyValMetProValValAsnValAspGlyTyTrhAspTySerTrpLysL 234
609 ..AATGATGTGAGAGAAAGACCGTTCTTTATGCGAAACATATGTCG 656
234 yAsnArGmetTrpArgLyAsnArGserPheTyTrhAlaAsnArHisCys 250
657 ATGGGAAGACSTGAATAGCACTTTGTCTGCAACACAGCGTGAGGA 706
251 ILeGlyThrAspLeuAsnArGAsnRheAlaSerLysHisTrpCysGluG 267
707 AGGTGCATCCAGTTCTCTGATGCTCGGAAACSTACTGTGACTTATCTG 756
267 uGlyAlaSerSerSerSerCysSerGluTrhTyTrpCysGlyLeuTyProG 284
757 AGTCAAGACCAAGAGTGAAGGAGCGTGTGTTCTTGAGAGAAATATC 806
284 IuSerIuProGluValValAlaValAlaLaserPheLeuArGAsnIle 300
807 AACCAATTAAGACATATACATGACATGATTCATCTCCAGCATATAGT 856
301 AsnGlnIleLeuAlaTyTrhIleSerMetHisSerTySerGlnHisIleVa 317
857 GTTTCATATCTCTTATACAGAAAGTAAAGCAAGACCATGAGAACTGT 906
317 IPrProTySerTyTrhArgSerLysSerLysAspHisGluGluLeuS 334
907 CTTCTAGTACGAGTGAAGAGTGTGCTATTTACAAAACATAATAAT 956
334 eTyLeuValAlaSerGluAlaValAlaArgAlaIleGluTyTrhSerLysAsn 350
957 ACCAGTATACACATGCGCATGCGTCAGAAACSTTATACCTAGCTCTG 1006
351 ThrArgTyTrhHisGlyHisGlySerGluTrhLeuTyLeuAlaProG 367
1007 AGTGGGAGCATGATGATGATGATTTGGGCATCAATATTCGTT..... 1051
367 yGlyGlyAspArTrpIleTyAspLeuGlyIleLysTySerPheThr 384
1052 .....TAC 1054
384 IeGluLeuArGAsnTrhGlyTrhTyTrhPheLeuLeuProGluArGTyTr 400
1055 ATCAAAACCACTGTAGAGAGCTTTGGCGGTCTCTAAATAGCTTG 1104
401 IeLysProThrCysArGArgGluAlaPheAlaIleValSerLysIleAlaTr 417
1105 GCATGTCATAGCAATGTT 1123
417 rHisValIleArGAsnVal 423
seq_name: sp_rident:Q9JNH6
seq_documentation_block:
ID Q9JNH6 PRELIMINARY; PRT; 422 AA.
AC Q9JNH6;
DT 01-OCT-2000 (TREMBLrel. 15, created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE CARBOXYPEPTIDASE R (THROMBIN-ACTIVATABLE FIBRINOLYSIS INHIBITOR)
DE (1110032P04RIK PROTEIN).
GN CPB2 OR TAFI OR 1110032P04RIK.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20341711; PubMed-10678383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not.";
RL J. Immunol. 165:1053-1058(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-20201996; PubMed-10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor.";
RL Thromb. Haemost. 83:297-303(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE-21085660; PubMed-11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Koehliwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudil F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzairelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AB021966; BAB03402.1; -
DR EMBL: AF164524; AAF62385.1; -
DR EMBL: AK004045; BAB23141.1; -
DR HSSP: P00730; 2CTC.
DR MEROPS: M14.009; -.
DR MGD: MGI:1891837; Cpb2.
DR InterPro: IPR003146; Proper_M14.
DR InterPro: IPR000834; Zn_carboxept.
DR Pfam: PF02244; Proper_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 1.
DR PRINTS: PR00765; CRBOXYPTSEA.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

```

alignment_scores:

```

Quality: 1592.50      length: 422
Ratio: 4.576          Gaps: 3
Percent Similarity: 82.464      Percent Identity: 73.460

```

alignment_block:

```

US-09-980-881-1 x Q9JNH6 ..
Align seg 1/1 to: Q9JNH6 from: 1 to: 422

```

```

18 ATGAACTTTGACGCTTGCAGCTTGTGACCACTGTTCTCTGTGTA 67
||||||| :|||:||||| |||:|||||
1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTy... 51 16
68 GCAGCATGCTTCGCGTCCAGAGTGGCAAGTCTAGCTGCTCTCCGA 117
||||||| ||||||| ||||||| ||||||| ||||||| |||||||
16 uGlnHisGlyPheAlaPheGlnSerIleGlnValLeuSerAlaLeuProA 33

```



```

118 GAACCTTAGCAAGTTCAAGTTTACAGAATCTTACTACAACTPAGAG 167
|||||
33 rgnrSerArgInValGlnLeuLeuGlnAsnLeuThrThrTyrGlu 49
|||||
168 ATTGTCTCGGCGGCGGTACAGCTGACCTTTGTGAGAAAGAAAACA 217
|||||
50 ValValLeuThrPglProValThrAlaGluPheIleGluLysLysG1 66
|||||
218 AGTCCATTTTGTGTAATGATCTGATGTCGACAAATGTGAAGCCCAT 267
|||||
66 uValHisPhePheValAsnAlaSerAspValAlaSerValLysAlaHisL 83
|||||
268 TAAATGTGACGGGAATTCATGACAGTGTCTTGCGACAGCTGGAAGAT 317
|||||
83 euAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99
|||||
318 CTATTATCAACAGCAGATTTCACAGACAGCTACGCCCGGAGCTCCGC 367
|||||
100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProAlaGlnAlaSerAl 116
|||||
368 ATGCTACTATGACAGCTATCATCTACTATAATGAATCTATTCTTGATAG 417
|||||
116 aserTyrTyrGluGlnIleThrHisSerLeuAsnGluIleTyrSerTrpIleG 133
|||||
418 AATTATTAAGTACAGAGCATCTGATATGCTTACAAAATCCACATTGGA 467
|||||
133 luValIleThrGlnGlnHisProAspMetLeuGlnLysIleTyrIleGly 149
|||||
468 TCCCATTTGAGAGTACCCACCTATGTTTAAAGTTTGTGAGAAAAGA 517
|||||
150 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysG1 166
|||||
518 ACAAAACAGCCAAAATGCCATATGATGATGATGATGATCCATCCAGAG 567
|||||
166 uGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgG 183
|||||
568 AATGATCTCTCTGCTTCTGCTTGCGTTCATAGGCCAT. .... 608
|||||
183 luTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199
|||||
608 ..... 608
200 PheHisGlyLysGlnAsnLeuTyrThrArgLeuLeuArgHisValAspPh 216
|||||
608 ..... 608
216 eTyrIleMetProValMetAsnValAspGlyTyrAspTyrThrTrpLysL 233
|||||
609 ..AATGCAATGTGGAGAAAGAACGGTTCTTTCTATGCGAACATCATTCG 656
|||||
233 yAsnAsnArgMetTrpArgLysAsnArgSerAlaHisLysAsnAsnArgLys 249
|||||
657 ATGCGAACAAGCCTGAATAGCACTTGTCTCCAAACACCTGCTGAGAGA 706
|||||
250 ValGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTyrCysGlyL 266
|||||
707 AGTGCATCTCAGTCTCATGCTCGAAACCTACTGTGAGACTTATCTGTG 756
|||||
266 sGlyAlaSerSerSerSerGluThrTyrCysGlyLeuTyrProG 283
|||||
757 AGTGAACAACAGAAAGTGAAGCAGTGGCTAGTTCTTGAGAAANAATC 806
|||||
283 luSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299
|||||
807 AACGAGATTAAAGCATATCAGCATGATCATCTCCAGCATATATGT 856
|||||
300 AspHisLysLeuAlaTyrIleSerMetHisSerLysSerGlnGlnIleLe 316
|||||
857 GTTTCATATTCCTATACACAGACAGTAAAGCAAGACCATGAGAACTGT 906
|||||
316 upPheProTyrSerTyrAsnArgSerLysSerLysAspHisGluGluLeu 333

```

```

907 CTCTAGTGACCAAGTACAGCAGTCTGCTATTTGACAAAACATGTAATAAT 956
|||||
333 eTleuValAlaSerGlnValAlaValArgAlaIleGluSerLysAsn 349
|||||
957 ACCAGTATACACATGAGCCATGGCTCAGAAACCTTATACCTAGCTCCG 1006
|||||
350 ThrArgTyrThrHisGlySerGlySerGluSerLeuTyrLeuAlaProG1 366
|||||
1007 AGTGGGAGCAGATTGATCTATGATTTGGCATCAAAATATTCGTT. .... 1051
|||||
366 yGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheTrpI 383
|||||
1052 .....TAC 1054.
|||||
383 leGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgTyr 399
|||||
1055 ATCAACCCACCTGTAGAGAGCTTTGGCCGCTGCTCTAAATATGCTTG 1104
|||||
400 IleLysProThrCysAlaGluAlaLeuAlaAlaIleSerLysIleValTr 416
|||||
1105 GCATGCTCATTAGAAT 1120
|||||
416 pHisValIleArgAsn 421
|||||

```

```

seq_name: sp_rodent:O9QZF0
seq_documentation_block:
ID O9QZF0 PRELIMINARY; PRT; 422 AA.
AC O9QZF0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE U.
GN CPB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CHAF1J;
RA He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
RT gene."
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF186188; AAF00528.1; -.
DR HSSP; P00730; SCPA.
DR MEROPS; M14.009; -.
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SEQUENCE 422 AA; 48832 MW; B53FFB09943954E CRC64;

```

```

alignment_scores:
Quality: 1586.50 Length: 422
Ratio: 4.572 Gaps: 3
Percent Similarity: 82.227 Percent Identity: 73.223

```

```

alignment_block:
US-09-980-881-1 x O9QZF0 ..

```

```

Align seg 1/1 to: O9QZF0 from: 1 to: 422

```

```

18 ATGAAGCTTTGAGGCTTGACGCTCTTATGACCAATGTTCTCTGTGA 67
|||||
1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr...G1 16
|||||
68 GCAGATGCTCTTGGCTTCCAGAGTGGCCAAATGTCTAGTGTCTTTCGA 117
|||||

```

```

16 uGlnHisGlyPheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProA 33
118 GAACCTTAGGCAAGTTCAAGTTCTTACAGAAATCTTACTACAACATATGAG 167
133 rGthSerArgGlnValGlnLeuGlnAlaSerLeuThrThrTrpGlu 49
168 ATTGTTCTGTGCAGCCGTAACAGCTGACCTTATTTGTGAGAAAAAACA 217
50 ValValLeuTrpGlnProValThrAlaGlnPheIleGluLysLysGlu 66
218 AGTCATTTTTTTTGAATGCAATGTCATGTCGCAATGTGAAGCCCAT 267
66 uValHisPhePheValAlaSerAlaSerAlaSerValLysAlaHis 83
268 TAAATGTGACGCGAATTCGACATGCTCTGTCGACAGCTGGAAGAT 317
83 euAsnValSerArgIleProPheAsnValLeuMetAsnValGluAsp 99
318 CTATATTCACAGACATTTCCAAAGCAGACAGTCAGCCCGAGCCCTCCG 367
100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAl 116
368 ATCGTACTATGAACAGATATCAGTCACTAAATGAATCTATCTTGGATAG 417
116 aSerTrpTrpGluGlnIleTrpHisSerLeuAsnGluIleTrpSerTrpIleG 133
418 AATTTATTAAGTGAAGGACATCTGATATGCTTACAAAATCCACATTTGA 467
133 lValLeuThrGlnGlnHisProAspMetLeuGlnLysIleTrpIleGlu 149
468 TCCCTATTTGAGAAAGTACCCACTCTATGTTTAAAGTTTCGGAAGAA 517
150 SerSerPheGluLysTrpLeuTrpValLeuLysValSerGlyLysGlu 166
518 ACAACAGCCAAATAATGCCATATGATGACTGTGATGATTCATGCCAGAG 567
166 uGlnArgIleLysAsnAlaIleTrpIleAspGlyIleHisAlaArg 183
568 AATGATCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
183 lUTrpIleSerProAlaPheCysLeuTrpPheIleGlyTrpValThrGln 199
608 ..... 608
200 PheHisGlyLysGluAsnLeuTrpThrArgLeuLeuAlaHisValAspRh 216
608 ..... 608
216 eTrpIleMetProValMetAlaSerValAspGlyTrpArgTrpTrpLys 233
609 ..AATCGAAATGTGAGAAAGAACGCTTCTTCTTATGCGAACAATCATTC 656
233 yAsnArgMetTrpArgLysAsnArgSerAlaHisLysAsnAspArgCys 249
657 ATCGGAACGACCTGAATAGCACTTGTCTCCAAACACTGGTGATGAGA 706
250 ValGlyTrpHisPheLysAsnArgAsnPheAlaSerLysHisIleTrpCysGlu 266
707 AGGTGCATTCAGTTCCTCATGCTCGGAAACCTACGTGACATTTATCTG 756
266 sGlyLysSerSerSerSerCysSerGluThrTrpCysGlyLeuTrpProG 283
757 AGTCAGAACGACAGTGAAGGAGTGGCTAGTTTCTTGAGAAAGAAATATC 806
283 lSerGluProGluValLysAlaValAlaAspPheLeuArgTrpAsnIle 299
807 AACGAGATTAAGCATATCATCATGATGATGATGATGATGATGATGATGAT 856
300 AsnHisIleLysAlaTrpIleSerMetHisSerTrpSerGlnGlnIleLe 316
857 GTTTCATTTCTTATACAGCAAGTAAAGCAAGACATGAGGAGACTGT 906
316 uPheTrpTrpSerTrpArgAsnArgSerLysSerLysAsnHisGluGluLeu 333

```

```

907 CTCTAGTACGCAAGTGAAGAGTTGCTGCTATTTGACAAAAGTAAATAAT 956
333 eLeuValAlaSerGluAlaValAlaArgAlaIleGlySerPheAsnLysAsn 349
957 ACCAGCTATACACATGCGCATGCGCTCAGAAACCTTACTACTACTCTG 1006
350 ThrArgTrpThrHisGlySerGlySerGluSerLeuTrpLeuAlaProG 366
1007 AGGTGGGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1051
366 yGlySerAspAspTrpIleTrpAspLeuGlyIleLysTrpSerPheThr 383
1052 .....TAC 1054
383 lGluLeuArgAspThrGlyArgTrpGlyPheLeuLeuProGluArgTrp 399
1055 ATCAAAACCCACCTGAGAGAGCTTTGGCGGCTGCTTAAATACTACTG 1104
400 lLeuLysProThrCysAlaGluAlaLeuAlaAlaIleSerLysIleValTr 416
1105 GCATGTCATTAGGAAT 1120
416 pHISValIleArgAsn 421

seq_name: sp_rodent:Q9EQV9

seq_documentation_block:
ID Q9EQV9 PRELIMINARY; PRT; 422 AA.
AC Q9EQV9:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PRE-PROCARBOXYPEPTIDASE R.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20471387; PubMed=11021404;
RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA Okada N., Okada H.,
RT "Molecular cloning and partial characterization of rat
RT procarboxypeptidase R and carboxypeptidase N.";
RL Microbiol. Immunol. 44:719-728(2000).
DR EMBL: AB042598; BAB18617.1; -.
DR HSSP: P00730; 2CTC.
DR InterPro: IPR003146; Proper_M14.
DR InterPro: IPR00634; Zn_carboxept.
DR Pfam: PF02244; Proper_M14; 1.
DR Pfam: PF00246; Zn_carboxept; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
SQ
SEQUENCE 422 AA: 48826 MW; FFFD32A51A9366C8 CRC64;

alignment_scores:
Quality: 1568.50 Length: 422
Ratio: 4.507 Gaps: 3
Percent Similarity: 82.464 Percent Identity: 72.275

alignment_block:
US-09-980-881-1 x Q9EQV9 ..
Align seg 1/1 to: Q9EQV9 from: 1 to: 422

18 ATGACATTTGACAGCCTTGACGCTTGTGACCATGTTCTCTCTGTGA 67
||||| ..... |||||
1 MetLysLeuTrpGlyLeuGlnValLeuValAlaIleLeuTrp...GI 16
68 GCAGCATGCTTGGCTCCAGAGTGGCCAGTTCTAGCGCTCTCCCTA 117
||||| ..... |||||

```

```

16 ulysHISGLyleuAlaPheGlnSerGlyHisValIleuSerAlaIleuPro 33
118 GAACCTTGAAGCAAGTTCAGTCTTACAGAAATCTTACTACACATATGAG 167
|||||
33 rghSerAlaGlnValGlnIleuLeuGlnAsnIleuThrThrTrpGlu 49
168 ATGTGTTCTGCGACCGCGTACAGCTGACCTTATGTGAGAAAAACA 217
|||||
50 ValIleuTrpGlnProValThrAlaGluPheIleGluIysIysGlu 66
218 AGTCATTTTTTTTGTAAATGCATGCTGATGTCGACAAATGTGAAGCCCAT 267
|||||
66 ValHisPhePheValAsnAlaSerAspValAsnSerValIysAlaTrp 83
268 TAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGCGAGACGTGAAGAT 317
|||||
83 euAsnAlaSerAlaGlnIleProPheAsnValIleuMetAsnValGluAsp 99
318 CTATTTCAACAGCAGATTTCCAAACGACAGTCAGCCCGGACCTCCGC 367
|||||
100 LeuIleGlnGlnIleThrSerAsnAspThrValSerProAlaIleSer 116
368 ATGCTACTATGACAGATTCATCCTCAATGAATGAAATCTATCTTGATAG 417
|||||
116 rSerTrpTrpGlnGlnIleThrHisSerLeuAsnGluIleTrpSerTrp 133
418 AATTTTAACTGAGAGGATCCTGATATGCTTACAAAATCCACATTTGA 467
|||||
133 lValAlleThrGlnGlnHisProAspMetLeuGlnIysIleTrpGly 149
468 TCCTCATTTTGAGAGTACCCCTCTATGTTTAAAGTTTCTGAGAAAGA 517
|||||
150 SerSerTrpGluIysTrpProLeuTrpValIleuIysValSerGlyIysG 166
518 ACAACAGCGCAAAATGCGCATATGATGATGTCGTGGAATCCATCCCGAG 567
|||||
166 uHisAlaGlnValAsnAlaIleThrPheAspCysGlyTrpHisAlaArg 183
568 AATGATCTCTCTGCTTCTGCTTGTGATGATAGCCAT..... 608
|||||
183 lTrpPheSerProAlaPheCysLeuThrPheIleGlyTrpValThrGln 199
608 ..... 608
200 PheHisGlyIysGluAsnThrTrpThrArgLeuLeuArgHisValAsp 216
608 ..... 608
216 eTrpTrpIleMetProValMetAsnValAspGlyTrpAspTrpThrTrp 233
609 ..AATGGAATGTGAGAAAGAACCGTTCTTCTATGCGAACAAATCATATGC 656
|||||
233 yAsnAsnArgMetTrpArgGlyAsnArgSerValHisMetAsnAsnArgCys 249
657 ATGGAAGACAGCTGAATAGCAACTTGTCTCAACACATGCTGTGAGGA 706
|||||
250 ValGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGlu 266
707 AGGTGATCCAGTTCCTCATGCTCGGAACCTACTGTGGAATTATGCTG 756
|||||
266 sGlyAlaSerSerPheSerCysSerGlnTrpCysGlyLeuTrpPro 283
757 AGTCAGAACCAAGTGAAGAGCAGTGGCTAGTCTTGAGAGAAATATC 806
|||||
283 lSerGlnProGluValIysAlaValAlaAspPheLeuArgArgAsnIle 299
807 AACCAATTTAAAGCATATCAGATGATGATCTATCTCCAGCATATAGT 856
|||||
300 AsnHisIleLeuAlaTrpIleSerMetHisSerTrpSerGlnIleIle 316
857 GTTTCATATTCCTATACAGAGTAAGAAAGAACCATGAGGAATGT 906
|||||
316 uPheProTrpSerTrpAsnArgSerLysSerLysAspHisGlnGluLeu 333

```

```

907 CTCTAGTACCGCAGTGAAGAGTTGCTGCTATGTGACAAACTAGTAAAT 956
|||||
333 eIleValAlaSerGluAlaValArgAlaIleGlnSerIleAsnIysAsn 349
957 ACCAGTATACACATGCGCATGCTGACAGAAACCTTATACCTAGTCTCG 1006
|||||
350 ThrArgTrpThrHisGlySerGlySerGlnSerIleuTrpLeuAlaPro 366
1007 AGTGGGAGCAGATGATGATCTATGATTTGGGCATCAATATTCGTTTCAT 1055
|||||
366 yGlySerAspAspTrpIleTrpAspLeuGlyIleLysTrpSerPheThr 383
1056 C..... 1057
383 lGluLeuArgAspThrGlyArgTrpGlyPheLeuLeuProGluArgPhe 399
1058 ..AACCCACCTTACAGAAAGCTTTGCGGCTGCTCTTAAATAGCTTG 1104
|||||
400 lIleLysProThrCysAlaGluAlaLeuAlaValSerIysIleAlaTr 416
1105 GCATGTCATTAGGAAT 1120
|||||
416 pHisValIleArgAsn 421

```

seq_name: sp_human:Q9NTI8

```

seq_documentation_block:
ID Q9NTI8 PRELIMINARY; PRT; 198 AA.
AC Q9NTI8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE BA139H14.2 (CARBOXYPEPTIDASE B2 (PLASMA)) (FRAGMENT).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tromans A.;
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137141; CAB92622.1; -.
DR HSSP; P00730; ICPX.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPRASEA.
KW Carboxypeptidase.
FT NON_TER 198
FT SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;

```

alignment_scores:

```

Quality: 1029.00 Length: 197
Ratio: 5.223 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.492

```

alignment_block:

US-09-980-881-1 x Q9NTI8 ..

Align seg 1/1 to: Q9NTI8 from: 1 to: 198

```

18 ATGAAGCTTTGACCGCTTGGACAGTCTTGRACCAATGTCTCTTCTGCA 67
|||||
1 MetLysLeuCysSerLeuAlaValIleuValProIleValIleuPheCysG 17
68 GCACATGTTCTTCCGTTCCAGAGTGGCCAGTCTGAGCTCTCTCTCTA 117
|||||
17 uGlnHisValaPheAlaPheGlnSerGlyGlnValIleuAlaIleuProA 34

```

```

118 GAACCTTAGCAAGTCAAGTCTACAGAACTTACTACACATATGAG 167
|||||
34 rghrSerArgInValGlnValLeuGlnAsnLeuThrTrpTyrGlu 50
168 ATTGTTCTGTGCAGCCGCTACAGCTGACCTATTTGTGAAGAAAACA 217
|||||
51 IleValLeuTrpGlnProValThrAlaAspLeuValLysLysGlu 67
218 AGTCATTTTGTGTAATGCATCTGATGTGACAAATGGAAGCCCAT 267
|||||
67 nvalHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
268 TAATATGAGCCGGAATTCATGCAGTGTCTGTGGCAGCGTGAAGAT 317
|||||
84 euAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAsp 100
318 CTATTCACAGCAGATTTCCACAGACACAGTACGCCCGCAGCCGCC 367
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaAlaSerAl 117
368 ATGCTACTATGAACAGTATCCTACTAAATGAAATCTATTTCTGTGATAG 417
|||||
117 aSerTyrTyrGlnGlnTyrHisSerLeuAsnGlnIleTyrSerTrpIleG 134
418 AATTATTAAGTACAGCATCCCTGATATGCTTCAAAAATCCATTTGGA 467
|||||
134 luhelIethrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
468 TCCTCATTTGAGAAGTACCCACTGTATGTTTAAAGTTTCTGAAAAGA 517
|||||
151 SerSerPheGlnLysTyrProLeuTyrValLeuLysValSerLysGlu 167
518 ACAACAGCCAAAATGCCATATGATGATGCTGGAATCCATGCCAGAG 567
|||||
167 uGlnAlaAlaLysAsnAlaIleThrIleAspCysGlyIleHisAlaArg 184
568 AATGATCTCTCCTGCTTCTGCTGTGTTGCTTACGCCAT 608
|||||
184 lUtrPileSerProAlaPheCysLeuThrPheIleGlyHis 197

```

seq_name: sp Vertebrate: 09PUF2

```

seq_documentation_block:
ID 09PUF2      PRELIMINARY;      PRT: 416 AA.
AC 09PUF2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE HOMOLOG.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Sclerozoa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT "Screening of Bothrops jararaca pancreas cDNA library.";
RL Submitted (Sep-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190274; AF01344.1; -.
DR HSSP; P09955; INSA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Proper_M14.
DR InterPro; IPR000834; Zn_cardopept.
DR Pfam; PF02244; Tropen_M14; 1.
DR Pfam; PF00246; Zn_cardopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ SEQUENCE 416 AA; 47723 MW; 4F99854DD72B7A7 CRC64;

```

alignment_scores:

```

Quality: 715.00      Length: 406
Ratio: 2.668      Gaps: 6
Percent Similarity: 66.010      Percent Identity: 37.192

alignment_block:
US-09-980-881-1 x 09PUF2  ..

Align seg 1/1 to: 09PUF2 from: 1 to: 416

60 TPCGTGAGCAGCATGCTTCGGCTCCAGAGTGGCCAACTTCTACTGC 109
|||||
14 PheAlaGlnThrThrValHisArgPheAspGlyGluLysValTyrArgVal 30
110 TCTTCTAGAACCTTAGCAAGTTCAGATCTTACGAATCTTACTACAA 159
|||||
30 lThrProArgAsnGlnLysProLysValTyrPheLeuAsnTyrLeuAlaAsn 47
160 CATATGAGATTTGTTCTGTGCGACCGGTAAACAGCTGATATGTGAAG 209
|||||
47 LeValGlnValAspPheTrpArgProAspSerValGluLeuValLysAla 63
210 AAAAAACAGTCCATTTTGTGAATGCATCTGATGTGACAAATGTGA 259
|||||
64 GluMetThrValAspPheArgIleGluAlaAspArgCysSerLysValGlu 80
260 AGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTGTGCGAGAG 309
|||||
80 uSerIleLeuGlnGlnSerGlyLeuAsnTyrGluIleLeuIleAspAsnL 97
310 TGGAAATCTTATTCACAGCAGATTTCCAGACAGCAGTCAAGCCCGCA 359
|||||
97 euGlnAlaValLeuAspArgGlnLeuAspAsnHisAla.....Arg 110
360 GCCTCCGCAATCGTACTATGAGAGTATCAGTACATGCTAATGAATCTATTC 409
|||||
111 ThrAlaGlyTyrAsnTyrGluLysTyrAsnSerTrpGluLysIleAspAl 127
410 TTGGATGAAATTTATACGTGAGAGCGATCTGATATGCTTACAAAATCC 459
|||||
127 aTrpThrAlaAspIleAlaAsnGlnLysProSerLeuValSerArgLeuG 144
460 ACATGGATCCCTCATTTGAGAGTACCCACTCATGTTTAAAGTTTCT 509
|||||
144 lIleIleTyrThrThrPheGlnGlyArgProMetCProLeuLysVal... 159
510 GGAAAAAGAACAAACAGCCAAATGCCATATGATGATGATGATGATGCA 559
|||||
160 GlyLysProGlyValAsnLysLysAlaIlePheIleAspCysGlyPheH 176
560 TGCCAGAGATGATGCTCTGCTGCTTCTGTGCTGTGCATV..... 602
|||||
176 salArgGlnTyrPileSerProAlaPheCysGlnTrpPheValArgGlu 193
602 ..... 602
193 lavalArgThrTyrGlyLysGluThrIleMetThrGlnLeuLeuAsnLys 209
602 ..... 602
210 LeuAspPheTyrIleLeuProValLeuAsnIleAspGlyTyrValTyrSe 226
603 ...GGCCATATTCGAATGTGAGAAAGAACCGTTCTTCTTATGGCAACA 648
|||||
226 rTrpLysGlnSerArgMetTrpArgLysThrArgSerValAsnAlaGlyS 243
649 ATCATTCATGTCGAGACAGACCTGAATAGCAATGTGTCTCCCAACACAG 698
|||||
243 eTrhCysAlleIleTyrAspProAsnArgAsnPhe...AspAlaAlaTrp 258
699 TGTGAGGAAGTGCATCCAGTCTCATGCTCGAAGCTGGAACCTACTGTGACT 748
|||||
259 CysSerValGlyAlaSerArgAsnProCysSerGlnTyrTrpCysGlySe 275

```

```

749 TTATCTGAGTCAGACAGACAGAGTGAAGCCAGTCAGTTCCTTGTGAGAA 798
    |||||
275 rlyspgcluserglulysglutlhrlyslalaLeuAlaAspHeileArgA 292
799 GAATATCAACAGATTAAAGCATATACATGATGATTCATCTACTCCAG 848
    |||||
292 rgaAnaYserllelelelnAlaTyrleuThrleHisSerTyrSerln 308
849 CATAATGTTTCCATATTCCTATACAGAAAGTAAAGCAAGACATGA 898
    |||||
309 MetleuLeuTyrProTyrSerTyrThrTyrAspLeuThrSerAsnAnly 325
899 GGAAGCTGTCTCTAGTACCCAGTGAAGAGTTCGCTGATTTGACAAACTA 948
    |||||
325 slyslAsnAsnSerlleAlaTylsGlulAlaIleArgGlulEu...lysValL 341
949 GTAAATAATACAGATATACATGCGCATGCGTCAAGAAAGCTTATACCTA 998
    |||||
341 eupheglYthrGluTyrThrTyrGlyProGlyAlaAlaThrIleTyrPro 357
999 GCTCCTGGAGGTGGGAGCATGATGATGATGATGATGATGATGATGATGAT 1048
    |||||
358 AlaAlaGlyGlySerAspAspTPrAlaTyrAspGlnGlyIleTyrAl 374
1049 GTTACATC..... 1057
374 aPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheAlaLeuProG 391
1058 .....AAACCCAGCTGTAGAGAAAGTTTGGCGTGTCTCTAAA 1096
    |||||
391 luserGlnIleLysProThrCysGluGlnThrMetIleAlaValLysTyr 407
1097 ATACCTGGCATGTCATT 1114
    |||||
408 lleAlaGluTyrMetleu 413

seq_name: sp_human:096E94

seq_documentation_block:
ID 096E94 PRELIMINARY; PRT; 417 AA.
AC 096E94;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SIMILAR TO CARBOXYPEPTIDASE A3 (MAST CELL).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW, AND ACUTE MYELOCYTIC LEUKEMIA;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
KW Carboxypeptidase.
KW SEQUENCE 417 AA; 48714 MW; F781EEF23F570E98 CRC64;

alignment_scores:
Quality: 652.50 Length: 427
Ratio: 2.417 Gaps: 10
Percent Similarity: 63.232 Percent Identity: 33.724

Alignment_block:
US-09-980-881-1 x 096E94
Align seg 1/1 to: 096E94 from: 1 to: 417
6 ATTCGTGTGGATGAAGCTTTGACGCTTCAGTCTGTACCATTTGT 55
::: |||||:::
5 leuProValGlyleuIleAlaThrThreuleAlaIle...AlaProVal.. 19

```

```

56 TCTCTTGTGTAGACAGCATGTCCTTCAGAGTGGCCAGTTCTAG 105
    |||||
20 .....ArgPheAspArgLysValPhe 28
106 CTGCTCTTCCTAGAAACCTTAGGCAAGTTCAAGTCTCAGAACTTCT 155
    |||||
28 rglValLysProGlnAspGlnLysGlnAlaAspIlelleLysPheuleA 44
156 ACAACATATGAGATTGTTCTGTGGAGCGGATGACAGTGCCTATATGT 205
    |||||
45 LysThrAsnGluLeuAspPheTyrProGlyAlaThrHisLysValAl 61
206 CAAGCAAAAACAGTCCATTTTGTGTAATGATCTGATGTCGCAATG 255
    |||||
61 aAlaLysMetMetValaAspPheArgValSerGluLysLysSerGlnAla 78
256 TGAAGGCCATTAAATGTAGCGGGAATTCATGATGAGTGTGCTGTGCA 305
    |||||
78 leGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleuIleHis 94
306 GACGTGAAGATCTTATTCACAGCAGATTC...AAGCAGACAGTCAG 352
    |||||
95 AspleuGlnGluGluIleGluLysGlnPheAspValLysGluAspIlePr 111
353 CCCCCGACCTCCGATGCTACTATGAACAGTATCAGTCACATAATGAA 402
    |||||
111 ogLysThrSer.....TyrAlaLysTyrAsnAsnTPrGluLysI 125
403 TCTATCTGTGATGATTAATTAAGTGAAGGAGCATCCGATATGCTTCA 452
    |||||
125 leValAlaTPrThrGluLysMetAspLysTyrProGluMetValSer 141
453 AAAATCCATTTGATCCTCATTTTGAGAGTACCCATCTATGTTTAA 502
    |||||
142 ArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuL 158
503 GGTTCGTGAAAGAACAAACAGCCAAAATGCCATATGATGATGATG 552
    |||||
158 sile...GlyLysAsnGluArgArgLysAlaIlePheMetAspCysG 174
553 GAATCATGCCAGAGATGATGATGATGATGATGATGATGATGATGAT 602
    |||||
174 lYlleHisAlaArgGluTyrPValSerProAlaPheCysGlnTPrPheVal 190
602 ..... 602
191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeu 207
603 .....GGCCAT. 608
207 uAspArgMetAsnPheTyrIleLeuProValPheAsnValAspGlyTyr 224
609 .....AATGSAATGTGAGAGAAAGACGTTCTTCAT 641
224 leTPrSerTPrThrLysAsnArgMetTPrArgLysAsnArgSerLysAsn 240
642 GCGAACAATTCATTTGCATCGAAGACAGCTGATACATTTGTTCCAA 691
    |||||
241 GlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe...AsnAl 256
692 ACAGTGTGTGAGAGAGTGCATGCTCATGCTGCTGCGAAGACCTACT 741
    |||||
256 aserTPrAsnSerlleProAsnThrAsnAspProCysAlaAspAsnTyrA 273
742 GTGACATTATCTGATGTCAGAACAGAGAGTGAAGGAGTGGCTGATTC 791
    |||||
273 rglYSerAlaProGluSerGluLysGluThrLysAlaValAlaThrAsnPe 289
792 TTGAGAAAGAAATATCAACAGATTAAAGCATATACATGATGATTCATA 841
    |||||
290 lleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyr 306
842 CTCGCCAGATATAGTGTTCATATTCCTATPACAGGAAGTAAAGCAAG 891

```

```

||||| 306 rserglnmetleuleupherprotlytyrthrserlysleupropiora 323
||||| 892 ACCATGAGCACTGTCTGTAGTACCAAGCACTGCTGTATTGAC 941
||||| 323 snhlsiglunspreualalysvalalalysileglythralspvalleu... 338
942 AAAACAGTAAATACACAGTATACATGGCCATGGCTCAGAAACCTT 991
339 SerThrArgTyrGlnThrArgTyrIleTyrGlyProIleGlnSerThrII 355
992 ATACCTAGCTCTGGAGTGGGAGCATTTGATCTATGATTTGGGCATCA 1041
355 eTyrProIleSerGlySerSerLeuasprrlAlaTyrAspLeuGlyIleI 372
1042 AATATTCGTTTAC..... 1054
372 yslHsrThrPheAlaPheGlnLeuAlaGspLysGlyLysPheGlyPheLeu 388
1055 .....ATCAAAACCCACTGTAGAGCACTTTGCGCGCTG 1089
389 leuprogIuSerArgIleLysProIleCysArgGluThrMetLeuAla 405
1090 CTCCTAAATAGCTTGGCAGTCATTAGCAAT 1120
405 llyspheilelAlaLysTyrIleLeuLysHis 415

seq_name: sp_human:Q96B08
seq_documentation_block:
AC Q96B08 PRELIMINARY; PRT: 417 AA.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEtical 47.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADRENAL CARCINOMA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC015338; AAI15338.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 417 AA: 47367 MW; EBBB9B27F5D5AF9 CRC64;

alignment_scores:
Quality: 653.50 Length: 422
Ratio: 2.394 Gaps: 8
Percent Similarity: 64.692 Percent Identity: 33.649

alignment_block:
US-09-980-881-1 x Q96B08 ..

Align seg 1/1 to: Q96B08 from: 1 to: 417

24 CTTTGCAGCTTGCACTGTGACCACTGTTCTTCTGTGAGCACA 73
::: ||||| ||||| ||||| ||||| |||||
1 MetleuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisH1 17
74 TGTCTTGGCG...TTCCAGATGGGCAAGTTTATGCTCTCTCTAGAA 120
| ||||| ||||| ||||| ||||| |||||
17 sGlyGlyGlnHisPheGlnGlyGlyValAlaPheArgValAlaHisGlnA 34
121 CTTTACGCAAGTCAAGTTCTACAGAACTTACTACACATAGAGATT 170
::: ||||| ||||| ||||| ||||| |||||
34 spGlnAsnHisIleAsnIleIleLeuArgIleLeuAlaSerThrThrGlnIle 50
171 GTTCTGTGGAGCGCGTAAACAGTACCTTATGTGAAGAAAAAAGT 220

```

```

||||| 51 AspPheTrpLysProAspSerValThrGlnIleLysProHisSerThrVa 67
||||| 221 CCATTTTGTAAATGCAATGCTGATGTCACAAATGGAAGGCCATTTA 270
| ||| ||||| ||||| ||||| ||||| |||||
67 lAspPheAlaValLysValAlaGlnAspThrValThrValGlnAsnValLeuII 84
271 ATGTGAGCGCAATTCATGACAGTGTCTTCTGCGACAGCTGGAGATCTT 320
::: ||||| ||||| ||||| ||||| |||||
84 yslGlnsnGlnLeuGlnIleTyrValLeuIleSerAsnLeuArgAsnVal 100
321 ATTCAAACAGCAGATTTCACACAGACAGTCAGCCCCGAGCTCCGCATC 370
||||| ||||| ||||| ||||| |||||
101 ValGlnAlaGlnPheAspSerArgVal.....ArgAlaThrGlnIleH1 114
371 GTACTATGAACAGATCATCAGTAAATGAAATCTATCTGATGAGCAAT 420
||||| ||||| ||||| ||||| |||||
114 sSerTyrGlnLysTyrAsnLysTrpGlnThrIleGlnAlaIlePrThrGlnG 131
421 TTATACTGAGAGCATCTGATATGCTTACAAAATTCACATTTGATTC 470
||||| ||||| ||||| ||||| |||||
131 lValAlaIleThrGlnAsnProAlaIleuIleSerArgSerValIleGlyThr 147
471 TCATTTGAGAAAGTACCCACTGTATGTTTAAAGTTCTTGGAAAGACA 520
::: ||||| ||||| ||||| ||||| |||||
148 ThrPheGlnGlyAlaGlnAlaIleTyrLeuLeuLysVal...GlyLysAlaG1 163
521 AACAGCCAAAATGCCATATGATGACTGTGTGCAATGCCACAGCAAT 570
||| ||||| ||||| ||||| ||||| |||||
163 yglAsnLysProAlaIlePheMetAspCysGlyPheHisAlaIleArgIle 180
571 GGAATCTCTGCTTCTGTCTGTTGGTCAFA..... 602
||||| ||||| ||||| ||||| |||||
180 rPleSerProAlaPheCysGlnIleTrpPheValArgGlnAlaValArgThr 196
602 ..... 602
197 TyrGlyArgGlnIleGlnValThrGlnLeuLeuAspLysLeuAspPheTy 213
603 .....GCCCAT.....A 609
213 rValLeuProValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysS 230
610 ATCGAATGTGGAGAAAGCAACGTTCTTATGCGAAACATTCATGATC 659
::: ||||| ||||| ||||| ||||| |||||
230 eArgPheTrpArgLysThrArgSerThrHisThrGlySerCysIle 246
660 GGAACAGACCTGAATGCAACTTGTCTCCAAACACTGATGTGAGGAAG 709
||||| ||||| ||||| ||||| |||||
247 GlyThrAspProAsnArgAsnPhe...AspAlaGlyTyrCysGlnIleG1 262
710 TGCATTCAGTTCCTCATGCTCGGAAACCTTACTGTGACTTATTCGTGACT 759
||||| ||||| ||||| ||||| |||||
262 yAlaSerArgAsnProCysAspGlnIleTyrCysGlyProAlaAlaIle 279
760 CAGAACCAAGATGAGCAAGTGGCTAGTTCTTGAGAGAAATATATCAAC 809
||||| ||||| ||||| ||||| |||||
279 eArgLysGlnThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295
810 CAGATTAAAGCATACATCAGATGCATTCATCTCCAGCATATAGTGT 859
::: ||||| ||||| ||||| ||||| |||||
296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerIleMetLeuTyr 312
860 TCCATATTCCTATACAGAGTAAACAAAGACCAATGAGCAACTGCTGC 909
||||| ||||| ||||| ||||| |||||
312 rProTyrSerTyrAlaTyrLysLysGlyGlnAsnAlaGlnLeuAsnA 329
910 TAGTAGCCAGTGAAGCACTTCTGTATGTGCAAAACTAGTAAATATCC 959
::: ||||| ||||| ||||| ||||| |||||
329 lAlaLeuAlaLysAlaThrValLysGlnLeu...AlaSerLeuHisGlyThr 344
960 AGGTATACACATGGCCATGGCTCAGAAACCTTATCTAGCTAGCTGGAG 1009
::: ||||| ||||| ||||| ||||| |||||

```


910 TAGTAGCAGTGAAGCAGTTCGTCTATTGACAAACTAGTAAATACC 959
 329 snleuAlaValAlaValIleu...AlaThrIleuTyrglyThr 344
 960 AGGTATACATGATGCGCTGCGTGAAGACCTTATACCTAGCTCTGAGG 1009
 345 LysThrThrTyrglyProGlyAlaThrThrIleuProAlaIleuIleu 361
 1010 TGGGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1055
 361 yserAspAspTrpAlaIleuValIleuValIleuValIleuValIleu 376

seq_name: sp.rodent:p97597

seq_documentation_block:

ID P97597 PRELIMINARY; PRT; 412 AA.

AC P97597;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MAST CELL CARBOXYPEPTIDASE A PRECURSOR (FRAGMENT).
 GN R-CPA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=SPRAGUE DAWLEY;
 RX MEDLINE=97149430; PubMed=8996238;
 RA Lutschschwab C., Pejler G., Aveskog M., Hellman L.;
 RT "Secretory granule proteases in rat mast cells. Cloning of 10
 RT different serine proteases and a carboxypeptidase A from various rat
 RT mast cell populations."
 RL J. Exp. Med. 185:13-29(1997).
 DR EMBL: U67914; AAB48267.1; -.
 DR HSSP: P09955; INSA.
 DR MEROPS: M14.010; -.
 DR InterPro: IPR003146; Proper_M14.
 DR InterPro: IPR000834; Zn_carboxpept.
 DR Pfam: PF02244; Proper_M14; 1.
 DR Pfam: PF00246; Zn_carboxpept. 1.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Signal; Carboxypeptidase.
 FT SIGNAL 1 1
 FT NON_TER 1 1
 FT SIGNAL 1 1
 FT CHAIN 11 412 MAST CELL CARBOXYPEPTIDASE A.
 SQ SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

alignment_scores:

Quality: 642.50 Length: 400
 Ratio: 2.530 Gaps: 7
 Percent Similarity: 63.500 Percent Identity: 35.750

alignment_block:

US-09-980-881-1 x P97597 ..

Align seg 1/1 to: P97597 from: 1 to: 412

84 TTCACAGATGCGCCAGTCTAGCTCTTCTAGAACCTTAGGCAAGT 133
 16 PheAspArgGluValPheArgValIleuGlnAspGluLysGlnAl 32
 134 TCAAGTCTACAGAACTTACTACAAACATATGATGATGATGATGATGAT 183
 32 AserIleuLysAsnLeuThrGlnThrIleuLysAspPheTrpTy 49
 184 CGGTACAGCTGACCTTATGTAAGAAACAAAGTCCATTTTGTGTA 233
 49 roAspAlaIleuIleuAspIleuAlaValAsnMetThrValAspPheArgVal 65

234 AATGATCTGATGTCGACAAATGTGAAGCCCATTTAAATGAGCGGAT 283
 66 ThrGluLysGluSerGlnThrIleuGlnSerThrLeuGlnIleuLysLys 82
 284 TCATGACAGTGTCTGCTGCGACAGCTGGAAGATCTTATCAACAGAGA 333
 82 LysTyrgluIleuLysAsnLysGlnGlnGlnGlnGlnGlnGlnGlnGln 99
 334 TTCACAGATGACAGTACGCGCCGAGCCCGCATCTGATGTAAGAG 383
 99 he.....AspValLysGluGluIleuAlaGlyArgHisSerThrAlaLys 113
 384 TATCACTCACTAAATGAATCTTATGATGATGATGATGATGATGATGAT 433
 114 TyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGlu 130
 434 GCATCTGATATGCTTACAAATTCACATGATGATGATGATGATGATGAT 483
 130 SHSPGluMetValSerArgIleuLysIleGlySerThrValGluAsp 147
 484 ACCACCTCTATGTTTAAAGCTTCTGGAAGAAAGAACAGCCAAAT 533
 147 snProLeuTyrgluLysIle...GlyArgLysAspGlyLysArgLys 162
 534 GCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 583
 163 AlaIlePheMetAspCysGlyIleHisAlaArgLysTrpAlaSerPro 179
 584 TTTCCTGCTGTGTTTAAAGCTTCTGGAAGAAAGAACAGCCAAAT 611
 179 aPheCysGlnTrpPheValTyrgluAlaLysSerTyrglyLysAsn 196
 611 611
 196 yslleuMetThrLysLeuLysAspArgMetAsnPheTyrgluLeuProVal 212
 612CGAATGTGAG 622
 213 PheAsnValAspGlyTyrlleTrpSerTrpThrLysAspArgMetTrp 229
 623 AAGAACCGCTCTTCTTATGCGCAACATCTTGCATGCAAGACCTGA 672
 229 GlyAsnAsnArgSerLysAsnProAsnSerThrCysIleGlyThrAspLeu 246
 673 ATAGCACTTGTGTCGCAAAACACTGCTGAGGAGAGGTGCATGACCTGC 722
 246 snArgAsnPhe...AspValSerTrpAspSerSerProAsnThrAspAsn 261
 723 TCATGCTCGGAACCTACTGTGACCTTTATCCTGATGACAGAACCAAGT 772
 262 ProCysLeuSerValTyrglyProAlaProGluSerGluLysGlu 278
 773 GAAGCAGTGGCTAGTCTTGTGAGAAATATCAACCGATTAAGCAT 822
 278 rLysAlaValThrAsnPheIleArgSerHisLysAsnSerIleLysAla 295
 823 ACATGACATGATCATCTATCTCCACATATAGTGTTCATATTCCTAT 872
 295 yrlleThrPheHisSerTyrglyMetLeuLeuPheProTyrglyTy 311
 873 ACAGAAAGTAAAGCAAGACCATGAGAACTGTCTTACAGCAGTGA 922
 312 ThrIleLysLeuProProAsnHisGlnAspLeuLysValAlaArg 328
 923 AGCAGTCTGCTATGACAAACTAGTAAATATCCAGCTTATACATG 972
 328 eAlaThrAspValLeu...SerSerArgTyrgluThrArgTyrlleTy 344
 973 GCCATGCTGCAAGAACTTATACCTAGCTCTGAGGAGGTGGAGCATGG 1022
 344 LyrProIleAlaSerThrIleTyrglyThrSerGlySerSerLeuAspTrp 360


```

1023 ATCTATGATTGGGCATCAATATTCGTTAC..... 1054
1055 .....ATCAACCCACCTGTA 1070
377 sgllySerGlyPheLeuLeuProGluSerThrIleLeuProThrCysL 394
1071 GAGAACTTTTGGCGCTCTCTAAATACCTTGCGATGATGAGAAAT 1120
394 ysglThrMetLeuSerVallyPheIleAlaIlyTyIleLeuIlyHis 410

seq_name: sp_rudent:Q9CVD1

seq_documentation_block:
ID Q9CVD1 PRELIMINARY; PRT; 279 AA.
AC Q9CVD1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 2210008M23R1K PROTEIN (FRAGMENT).
DE 2210008M23R1K.
GN 2210008M23R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57Bl/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawal J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giusti C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Sakai K., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsi G.,
RA Blake J., Boftell D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RA *Functional annotation of a full-length mouse cDNA collection.*;
RT Nature 409:685-690(2001).
RL EMBL; AK008678; BAB25826.1; -.
DR HSSP; P09955; INSA.
DR MGD; MGI:1923953; 2210008M23R1K.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CRBOXYPTASE.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; UNKNOWN_1.
DR NON_TER 279
SQ SEQUENCE 279 AA; 31970 MW; 29468AFA0B67FOAA CRC64;

alignment_scores:
Quality: 531.00 Length: 226
Ratio: 3.493 Gaps: 4
Percent Similarity: 67.257 Percent Identity: 45.575

alignment_block:
US-09-980-881-1 x Q9CVD1 ..
Align seg 1/1 to: Q9CVD1 from: 1 to: 279
489 CTCATGTTTAAAGGTTTGTGAAGAACAAACAGCCAAATATGCAT 538
:::|||||::: |||||::: || |||:::

```

```

1 MetlyrValleuIlySile...GlyIySAspArgProAsnIlyProAlaPh 16
539 ATGATTCAGTCTGGATATCATGCCAGAAATGATCTCTCCGCTTCT 588
:::|||||:::|||||:::|||||:::|||||:::
16 ephellleAspCysGlyPheHisAlaArgLutPrIleSerProAlaPheC 33
569 GCTTGTTGTTGTTA..... 602
33 ysglThrPheValArgLutAlaValArgThrTyIyGlnGluIleHis 49
602 ..... 602
50 MetlyrArgLeuLeuAspGluLeuAspPheTyrValleuProValAs 66
603 .....GCCATATGCAATGTGAGAAAGA 627
66 nileAspGlyTyrValIlyThrThrTrpAlaIlySAspArgMetThrArgLysT 83
628 ACCGTTCTTTCTATGCAACATCATTCGATCGCAACAGACCTGAAATAGC 677
:::|||||:::|||||:::|||||:::|||||:::
83 hrArgSerThrThrAlaGlySerSerCysPheGlyValAspProAsnArg 99
678 AACTTGTCTCCAAACACTGCGTGAGAGAGTGATGATGATGATGATG 727
|||||:::|||||:::|||||:::|||||:::|||||:::
100 AsnPhe...AspAlaGlyTrpCysGluValGlyAlaSerArgSerProCys 115
728 CTCGGAACCTTACTGCTGATTCCTGAGTACAGACAGCAAGTGAAG 777
|||||:::|||||:::|||||:::|||||:::|||||:::
115 sserAspThrTyIyCysGlyProThrProGluSerGluIySglutThrIlySA 132
778 CAGTGGCTAGTTCTTGAGAGAAATATCAACAGATTAAAGCATATC 827
|||||:::|||||:::|||||:::|||||:::|||||:::
132 lAlaValAspPheIleArgLutGlnAsnLeuSerIleIlySAlaTyIyLeu 148
828 AGCAGATTCATCTACCCAGATATGATGATGATGATGATGATGATG 877
|||||:::|||||:::|||||:::|||||:::|||||:::
149 ThrValHisSerTyIySerGlnMetIleuTyIyProTyIySerTyIyAspTy 165
878 AAGTAAAGCAACAGCATGAGAACTGCTCTAGTACGACCTGAGCAG 927
:::|||||:::|||||:::|||||:::|||||:::|||||:::
165 rlySerLeuProGluAsnTyIyGlnIleuAsnAlaIleuVallySglutAla 182
928 TTCGCTGATTCAGCAAACTAGTAAATATCCAGATATACATGCGCAT 977
:::|||||:::|||||:::|||||:::|||||:::|||||:::
182 lAlySglutLeu...SerThrLeuHisGlyThrIlyTyIyThrTyIyPro 197
978 GCGCTGAGAACTTATACCTGATGCTGAGGATGAGGATGATGATG 1027
|||||:::|||||:::|||||:::|||||:::|||||:::
198 GlyAlaThrThrIleTyIyProAlaAlaGlyIySerAspSerPalaty 214
1028 TGATTTGGGCATCAATATTCGTTTACA 1055
||| |||||:::|||||:::|||||:::
214 rAspIlyGlyIleTyIySerPheThr 223

seq_name: sp_mammal:Q9TV85

seq_documentation_block:
ID Q9TV85 PRELIMINARY; PRT; 419 AA.
AC Q9TV85;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE A1 (BC 3.4.17.1).
GN PCPAl.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99192816; PubMed=10092856;
RX Darnis S., Juge N., Marino C., Aviles F.X., Pulgserver A., Chalik J.C.,
RA Guo X.J.;

```

"Cloning, sequencing and functional expression of a cDNA encoding
 RT porcine pancreatic preprocarboxypeptidase A1."
 Eur. J. Biochem. 259:719-725(1999).
 DR EMBL: AF076222; AAD17690.1; -
 DR HSSP: P09954; 1PCA.
 DR MEROPS: M14.001; -
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_CarDopept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_CarDopept; 1.
 DR PRINTS: PR00765; CRBOXYPTASE.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 DR HydroLase; Carboxypeptidase.
 KW SEQUENCE 419 AA; 47235 MW; 8484CB57B714FC1 CRC64;

alignment_scores:
 Quality: 516.50 Length: 391
 Ratio: 2.207 Gaps: 12
 Percent Similarity: 59.847 Percent Identity: 33.760

alignment_block:
 US-980-881-1 x 09TV85 ..

Align seg 1/1 to: 09TV85 from: 1 to: 419

```

24 CTTTGCAGCCTTGCAAGTCTGTACAGATCTTCTCTCTGTGAGCACA 73
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
4 LeuLeuIlePheSerValLeuLeuGlyGlyValLeuAlaLysGluAsp.. 19
74 TGTCTCCGCTCCAGAGTGGCCAGTCTTACGCGCTCTCTCTGAAACCT 123
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
20 .....PheValGlyHisGlnValLeuArgIleSerValAspArg 33
124 CTAGGCAAGTTCAAGTTCTACAGATCTTCTACACATATGAGATGTT 173
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
33 IuAlaGlnValGlnLysValLysGlnLeuGlnLysPheGlnHisLysGln 49
174 CTC.....TGGCAG...CCGGTAAACAGTGCACCTTATGTGAAGAAA 214
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
50 LeuAspPheThrPargIleProAlaArgProGlyPheProIleAspVal 66
215 ACAAGTCCATTTTGTCTAAATGATCATGTATGTCAGCAATGTGAAGCCC 264
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
66 9.....ValProPheProSerIleGlnAlaValLysValP 78
265 ATTGTAATGTGAGCGGAATTCATGCAAGTCTTGTGTCGACAGCTGAA 314
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
78 heLeuGlnAlaHisGlyIleArgTyrThrIleMetIleGlnAspValGln 94
315 GATCTTATTT.....CAACAGCATTTTCCAGACGACAGCTGACGCC 355
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
95 LeuLeuLeuAspArgIleGlnGlnGlnIleMetPheAlaSerGlnIleArg 111
356 CCGAGCCTCCGCACTCTGAC...TATGACAGATTCATCTACTCTAAAGAAA 402
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
111 ValGlyThrIleSerThrPheAsnTyrAlaThrTyrHisThrLysGlnGln 128
403 TCTATCTTGATGATAGATTATTAACAGAGAGCATCTGATATGCTTACA 452
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
128 IeTyrAspPheMetAspIleLeuValAlaGlnHisProGlnLeuValSer 144
453 AAAATCCATATGGATCCATCTTGTGAGAGTACCCCATCTATGTTTAA 502
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
145 LysLeuGlnIleIleYserSerTyrGlnIleArgProIleTyrValLeu 161
503 GCTTCTCGAAAAAGAACACCCAAAAT.....GCCATATGGA 543
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
161 sPheSer.....ThrGlyLysHisAsnArgProAlaIleTyrI 174
544 TTGACTGTGGAATCCATGCCAGAGAAATGATCTCTCTGCTTCTGCTTG 593
   ||| :|||:|||||:|||||:|||||:|||||:|||||:

```

```

174 IeAspThrGlyIleHisSerArgGlnIleValThrGlnAlaSerGlyVal 190
594 TGGTTCATTA..... 602
191 TrpPheAlaLysLysIleThrGlnAspTyrGlyGlnAspProAlaPheTh 207
602 ..... 602
207 AlaIleLeuAspHisLeuAspIlePheLeuGlnIleValThrAsnPro 224
603 .....GCCATATGCAATGTGAGAGAACCCGT 632
224 spGlyPheAlaPheThrHisSerGlnAsnArgMetThrPargLysThrArg 240
633 TCTTTCTATGAGAACATCATTCATGCGAAGACGCTGATGCAACT 682
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
241 SerArgThrSerGlySerPheCysValGlyValAspProAsnArgAsnTr 257
683 TGTCTCCAAACACTGTGTGAGAGAGTGCATCCAGTCCCTCATGCTCG 732
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
257 P...AspAlaGlyPheGlyGlyAlaGlyAlaSerSerAsnProCysSer 273
733 AAACCTACTGTGACTTATCTGAGTCAGACACAGAGTGAAGCAGTG 782
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
273 IuThrTyrHisGlyLysPheProAsnSerGlnValGlnValLysSerIle 289
783 GCTAGTTCTTGAGAAAGAAATATCAACGATTAAGCATATCAAGCAT 832
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
290 ValAspPheValAsnAspHisGlyAsn...IleLysAlaPheIleSerI 305
833 GCATTCATCTCCACAGATTAAGTGTCTTCATATTCCTATACAGAA 882
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
305 eHisSerTyrSerGlnLeuLeuLeuTyrProTyrGlyTyrLysThrGln 932
883 AAGCAAGAACATGAGAACTCTCTCTAGTACCAAGTGAAGATTCGT 932
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
322 IeProAlaAspLysAspGlnLeuAspGlnIleSerLysSerAlaValAla 938
933 GCTATTGACAAACTAGTAAATATACAGGATATACATGAGCATGGCTC 982
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
339 AlaLeu...ThrSerLeuTyrGlyThrLysPheGlnTyrGlySerIle 354
983 AGAAACCTTATACCTAGCTCTCTGAGGTGGGAGCATGATATGAT 1032
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
354 eThrThrIleTyrGlnAlaSerGlyGlyThrIleAspTrpThrTyrAsn 371
1033 TGGCATCAAAATATGCTTTACA 1055
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
371 IuGlyIleLysTyrSerPheSer 378

```

seq_name: sp_human:0960N3

seq_documentation_block:

```

ID 0960N3 PRELIMINARY; PRT; 417 AA.
AC 0960N3:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBOXYPEPTIDASE A2 (PANCREATIC).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteleia; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007009; AA07009.1; -
SEQUENCE 417 AA; 46756 MW; D1886784DFA862 CRC64;

```

alignment_scores:

Quality: 511.00 Length: 394
 Ratio: 2.156 Gaps: 11
 Percent Similarity: 60.152 Percent Identity: 32.487

alignment block:
 US-09-980-881-1 x Q96QN3 ..

Align seg 1/1 to: Q96QN3 from: 1 to: 417

```

72  CANGTCCTC.....GGCTCCAGATGGCCAGTTCCTAGCTGCTCT 112
    |||:::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
13  HistIeYrCysLeuGIuThrPheValGIYAspIuValLeuGIuIleVa 29
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
113 TCCTAGAACCTCTAGGCAAGTTCAGATTCTACAGAACTTACTACAAAT 162
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
29  lProSeRAsnGIuNGIuNGIleYAsnLeuGIuNGIuNGIuNGIaNG 46
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
163 ATGAGATTGTTCTC.....TGGCAGCGGTAACAGCTGACCTTATGTG 206
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
46  luhIstIeUGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGI 58
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
207 AAGAAAAAACAGATCCATTTTGTGTAATGATCATGTGATGCAAAATG 256
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
59  ProGIuYrIuThrAlaHisValArgValIProPheValAsnValGIuAlaVa 75
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
257 GAAAGCCCATTTAAATAGTGAAGCGGAATTCATGCAAGTCTTGCTGGCAG 306
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
75  lYsValrPheLeuGIYserGIuNGIYleAlaYrSerIleMeTIleGIuA 92
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
307 ACGTGAAGATCTTATTCACAGCAGATTCACAGCAGACAGCTGACGCC 356
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
92  spValGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 108
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
357 CGAGCCTCCGCATCG.....TACTATGACAGTATCACTCACTAA 397
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
109 ArgYrGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 125
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
398 TGAATCTATTCCTTGATACATTTTAACTGAGAGGCATCTGATATGC 447
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
125 uGIuIleSeRGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 142
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
448 TTACAAAAATCCAGATTCGATCCCTATTTGAGAGTACCCACTATGTT 497
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
142 alSeRlYsValAsnIleGIYseRseRpheGIuNGIuNGIuNGIuNGI 158
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
498 TTAAGGTTCT...GGAAGAAAGCAAAACAGCCAAAAATGCCATATGAT 544
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
159 leuYsPheSeRthrGIYGIYAsP.....LYSProAlaIleTrrple 172
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
545 TGAAGTGAATCCAGCCAGAGATGATCTCTGCTGCTTGTGCTTGT 594
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
172 uAspAlaGIYleHisAlaArgIuNGIuNGIuNGIuNGIuNGIuNGIuNG 189
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
595 GGTTCATAGGC..... 605
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
189 rPhIuAlaAsnLYsIleValSeRAsPTrGIYLYsAsPProSeRleThr 205
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
605 ..... 605
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
206 SerIleleuAspAlaLeuAspIlePheleuNGIuNGIuNGIuNGIuNG 222
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
606 .....CATTAATCGAATGTGAGAAAGACCGTT 633
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
222 pGIYTYrValPheSeRGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 239
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
634 CTTTCTATGGAGAACATCATGGAGAACAGACCGAATAGCAACTTT 683
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
239 erYsValSeRGIYSerleuCYsValGIYValAsPProAsnArgAsnTrp 255
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
684 GTCTCCAAGACCTGTGTGTGAGGAAGTGCATCCAGTTCCTCATGTCGA 733
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
256 ...AspAlaGIYrPheGIYGIYProGIYAlaSeRseRAsnProCYsSeRAs 271
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|

```

```

734 AACCTACTGTGACTTATTCCTGAGTCAAGAACAGAACTGAAGGCATGG 783
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
271 pSeRTrHisGIYProSeRAlaAsnSeRGIuNGIuNGIuNGIuNGIuNG 288
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
784 CTAGTTTCTTGGAAGAAATATCAACAGATTAAAGATACATCAGCATG 833
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
288 alAsPpheIleYsSeRHis...GIYLYsValLYsAlaPheIleThrleu 303
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
834 CATTCATACCTCCAGCATATAGTGTGTTCCATATTCCTATACAGAACTA 883
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
304 HisSeRTrYseRGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 320
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
884 AAGCAAAACCATGACGAGACTGTCTCTAGTACCCAGTGAAGCAGTCTGT 933
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
320 sLeuAsPAsPpheAsPGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 337
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
934 CTATGACAAAACATAGTAAATAATACAGGTATACATAGGCCATAGCCCA 983
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
337 erleu...ArgSeRleuHisGIYrHisTrYrLYsValGIYProIleCYs 352
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
984 GAAACCTTATACCTAGCTCTGAGAGTGGGAGCAGATTGATCTATGATTT 1033
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
353 SeRValIleTrYrGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 369
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
1034 GGCATCAAAATATTCGTTTACATCAAC..... 1061
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
369 rGIYleYTrYseRpheAlaPheGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 386
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
1062 .....CCACCTGTAGAAAGCTTTGGCG 1085
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
386 lYpHeleuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNGIuNG 396
    ||| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|

```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2002, 15:18:17 ; Search time 2123.6 Seconds

(without alignments)
15500.773 Million cell updates/sec

Title: US-09-980-881-1

Perfect score: 1573
Sequence: 1 agaaaattcgtgtggatg.....aaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Length	ID	Description

1	1528	97.1	1546	9	AB011969	AB011969 Homo sapi
2	1372	87.2	1728	9	HUMPCPBX	M75106 Human prepr
3	1372	87.2	1749	6	I16100	I16100 Sequence 2
4	1372	87.2	1749	6	I33526	I33526 Sequence 2
5	1309	83.2	1715	9	BC007057	BC007057 Homo sapi
6	918	58.4	1272	6	AR086324	AR086324 Sequence
7	656	41.7	1421	10	AB021968	AB021968 Mus muscu
8	656	41.7	1490	10	AF164524	AF164524 Mus muscu
9	654	41.6	1430	10	AF186188	AF186188 Mus muscu
10	652.6	41.5	1425	10	AB042598	AB042598 Rattus no
11	475.4	30.2	58097	9	AL157758	AL157758 Human DNA
12	409	26.0	437	11	G05971	G05971 human STS W
13	173.6	11.0	181	11	G59992	G59992 SHGC-130613
14	128	8.1	180680	9	AL137141	AL137141 Human DNA
15	124	7.9	927	5	AR072911	AR072911 Sequence
16	123.8	7.9	1327	5	AF190274	AF190274 Bothrops
17	123.4	7.8	921	6	I67699	I67699 Sequence 3
18	123.4	7.8	1215	6	I67698	I67698 Sequence 1
19	123.4	7.8	1251	4	SSC13375	AJ133755 Sus scrof
20	123.2	7.8	451	11	G53326	G53326 SHGC-82320
21	120.8	7.7	1251	4	DOCZAP47	D78348 Dog RNA fo
22	109	6.9	1263	6	AB7530	AB7530 Sequence 11
23	107.6	6.8	1622	9	HUMCARWC	M27717 Human mast
24	107.6	6.8	1622	11	G28614	G28614 human STS S
25	107.6	6.8	1674	9	BC012613	BC012613 Homo sapi
26	107.4	6.8	999	6	AS1908	AS1908 Sequence 72
27	107.4	6.8	999	6	AR085871	AR085871 Sequence
28	107.4	6.8	1053	6	AS1904	AS1904 Sequence 68
29	107.4	6.8	1053	6	AR085868	AR085868 Sequence
30	107.4	6.8	1263	6	AS1896	AS1896 Sequence 60
31	107.4	6.8	1263	6	AR085861	AR085861 Sequence
32	107.4	6.8	1284	6	AS1913	AS1913 Sequence 77
33	107.4	6.8	1284	6	AR085875	AR085875 Sequence
34	107.4	6.8	1310	9	HS2424866	AJ224866 Homo sapi
35	107.4	6.8	1389	9	BC015104	BC015104 Homo sapi
36	107.4	6.8	1454	9	BC015338	BC015338 Homo sapi
37	104.2	6.6	1053	6	AX000417	AX000417 Sequence
38	104.2	6.6	1053	6	AX000418	AX000418 Sequence
39	104.2	6.6	1059	6	AS1917	AS1917 Sequence 81
40	104.2	6.6	1059	6	AS1921	AS1921 Sequence 85
41	104.2	6.6	1059	6	AR085878	AR085878 Sequence
42	104.2	6.6	1059	6	AR085881	AR085881 Sequence
43	102.8	6.5	927	9	S40234	S40234 mast cell c
44	101	6.4	1870	6	A67356	A67356 Sequence 11
45	101	6.4	1870	6	A87561	A87561 Sequence 42

ALIGNMENTS

RESULT	1	1546 bp	mRNA	linear	PRI 02-FEB-2000
AB011969	AB011969	Homo sapiens mRNA for carboxypeptidase B-like protein, complete cds.			
LOCUS	AB011969				
DEFINITION	Homo sapiens mRNA for carboxypeptidase B-like protein, complete cds.				
ACCESSION	AB011969	GI:6855463			
VERSION	AB011969.1	GI:6855463			
KEYWORDS	carboxypeptidase B-like protein.				
SOURCE	Homo sapiens CDNA to mRNA.				
ORGANISM	Homo sapiens				
REFERENCE	Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.				
AUTHORS	1 (sites)				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
JOURNAL	Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma				
MEDLINE	J. Biol. Chem. 266 (32), 21833-21838 (1991)				
REFERENCE	92042093				
AUTHORS	2 (sites)				
TITLE	Matsumoto, A.				
JOURNAL	Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma				
REFERENCE	Unpublished (1998)				

KEYWORDS plasma carboxypeptidase.
SOURCE Homo sapiens liver CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1728)
AUTHORS Eaton, D.L., Malloy, B.E., Tsai, S.P., Henzel, W. and Drayna, D.
TITLE Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
JOURNAL J. Biol. Chem. 266 (32), 21833-21838 (1991)
MEDLINE 92042093
FEATURES
source location/Qualifiers
1..1728 /organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
1..1728 /gene="PCPB"
/gene="PCPB"
20..1291 /note="prepro-plasma carboxypeptidase B"
/codon_start=1 /protein_id="AA60042.1"
/db_xref="GI:189687"
/translation="MKLCSTAVLPYVIFCEQHFAPSGQVLAALPRTSKQVYLON
LITTEIVLMQPTADLVKQYHEFVNADVDVKNALHNSGIPCSVLADVEDLI
QOQISNDEIVSPRASASYEDYHSLNEIYSWLEFTEERDMVTKIHSSPEKPLVY
LKVSGKEQYKNAIWDICGIDHAREMISPAFLMFIHTOFQGIIGQYTNLRLVDFY
LMPVNDGYDYSWKNNMWRKNSFYNNHCIGDILRNFSKHWCEGASSSCSE
TYGGLYPESEPEKVAVALFLRNINQIKAYISMSISQHIYIPVSYTSKSKDHELS
LVSEAVRAIKETSKNRTYTHGSEETIYLAPEGGDDMIYDLGIRSYTIELRDTGTY
GFLPERIAFKTEKAEFAVSKIAHVIKRN"
mat_peptide /gene="PCPB"
86..1288 /product="plasma carboxypeptidase B"
BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN

Query Match 87.2% Score 1372; DB 9; Length 1728;
Best Local Similarity 90.3% Pred. No. 4.3e-307;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

QY 1 agaaaattgctgttggaagaagcttgcagccttgagctctgttaaccattgtctct 60
Db 3 AGAAAATTGCTGTTGGTGAAGCTTTGGACGCTTGACCTTGACCCATTGTTCTCT 62
QY 61 tctgtgaagcaagatcttgcgcttcagagtgagcaagtttagctgtctctccagaa 120
Db 63 TCTGTGAGCAGCATGTCTGCGGTTTCAGAGTGCCCAAGTTCTAGCTGCTTCCTAGAA 122
QY 121 cctctaaggaagttcaagttctacagaatcttactacaacataagatgttctctgcg 180
Db 123 CCTTAGGCAAGTTCAAGTTCTACAGAACTTACTACAAATATGATGTTCTCTGCG 182
QY 131 agcgggtacagctgacattatgtgaaagaaacaaagtcacatttttctgtaaatgat 240
Db 133 AGCCCGTAAACAGCTGACCTTATTTGGAAGAAAAACAAGTCCATTGTTTGGTAAATGCAT 242
QY 241 cgtatgtcacaaatgtgaaagccatttaaatgtgagcggaattccatcagatgtcttc 300
Db 243 CTGATGTGCACATGTGAAGCCCATTTAAATGTGAGCGGAATTCATCAGATGCTTGC 302
QY 301 tggcagagctggaagatcttatacagaagcagattccacagacacagcagcccccag 360
Db 303 TGGCAGAGCTGGAAGATCTTATTCAACAGCAGATTTCCAAAGACACAGACGCCCGAG 362
QY 361 cctcgcagctgtaactatgaacagatcactcaactaaatgaatctatcttggatagaat 420
Db 363 CCTCGCATCGTACATGAAAGATATCACTCACTAAATGAATGTAATCTTGATGAAAT 422
QY 421 ttatacgaagagcattcctgatactgatacaaaaatccacattgatactcatttgaaga 480

Db 423 TTATACTGAGAGGACATCCTGATATGCTTACAAAAATCCACATTTGATCCTCATTTGAGA 482
QY 481 agtaccactctatgttttaaaagttctcgtaaaagaacaaacagcaaaatgcatat 540
Db 483 AGTACCACCTCATATGTTTAAAGGTTTGGAAAAAACAACAGCCAAAAATGCCATAT 542
QY 541 ggattgaactgtgaaatccatgcagaagaatgagatctctctctctctctgttggttca 600
Db 543 GGAATTACGTGGAAATCCATGACAGAGAAATGATCTCTCGCTTCTGCTGCTGCTCA 602
QY 601 taggcacat----- 608
Db 603 TAGGCCATATTAACCTAATTCATGGAATTAATAGGCCAATATACCAATCTCTGAGCCTTG 662
QY 609 -----a 609
Db 663 TGGATTTCATGTTATGCGCGGTGTTAATGTGACGCTTATGACTCTCATGAAAAAGA 722
QY 610 atcgaatgtgagaagaacagcttctctctatgagacaatcatgcatcagaagacc 669
Db 723 ATCGAATGTGAGAAAGAACCGCTTCTTCTATGCGAACAATCATTCATCGAAGACACC 782
QY 670 tgaatgcaactgtctctcaaacactgtgtgtgagaagaagtgatccagttctctatgct 729
Db 783 TGAATAGCAACTTGTCTCCAAACACTGATGATGAGAGGTGATCCAGTCTCATGCT 842
QY 730 cggaaacctactgtgacttatacctcgtatcagaaccaaagtgaaagcagtgactagt 789
Db 843 CGGAACCTACAGTGTGACCTTATCCGATGACAGAACAGAGGAAGGACAGTGTGCTAGTT 902
QY 790 tcttgagaagaatcacaacagatlaaagcatatcatcagcatgcatcactccagc 849
Db 903 TCTTGAGAAAGAAATATCAACACAGATTAAGACATACATCAGCATCATCTATCTCCACG 962
QY 850 atatagtgttccatrttctctatcacagaagtaaaagcaagaacccatggagactgttc 909
Db 963 ATATAGTGTTCATATTCCTATACAGAGTAAGAAAGCAACGATGAGAACTGTCTC 1022
QY 910 tagtaagcagtgaaagcagtcgtctatctgacaaactagtaaaatacagaagtatacac 969
Db 1023 TAGTACCAGTGAAGCAGTTCGTGCTATTTGAGAAACTGTAAATATACAGGTATACAC 1082
QY 970 atggcagtgctcagaacattatcctagctctcgtgaaagtgggaagattgagatcatg 1029
Db 1083 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGGACGATTTGGATCTAG 1142
QY 1030 atttggagcacaatatgt----- 1050
Db 1143 ATTGGGATCAAAATATTTGCTTACAAATTGAACTTGAGATACGGGACATACGATTTCT 1202
QY 1051 -----ttacataaacccacactgttagagaagcttttgcgctgtctctaaaa 1097
Db 1203 TGTGTCGGGAGCGTTACATCAAAACCCAGCTGTAGAGAAAGCTTTTGGCGCTGTCTTAAA 1262
QY 1098 taagcttgcatgtcatatgaagatgtttaaagccctgatttatacattctgtctccgat 1157
Db 1263 TAGCTTGGCATGTCAATAGGAATGTTAAATGCCCTGATTTTATTCATGTGCTCCGATAT 1322
QY 1158 tttaattactgaattccagaagaacaaatcatatgataagatatttttaagttttac 1217
Db 1323 TTTAAATTAAGTACTGATTCAGAGAACCAAAATCATGTATCATGATTAATTTTAACTTTATC 1382
QY 1218 cgtagtttgaataaagatttccctatctcctgtgtctgttcagagaacctaataagtgtc 1277
Db 1383 CGTAGTTTGAATPAAAGATTTTCTATTCCTTGTTCTGTGCANAGAACCTAATAAGTGTCT 1442
QY 1278 acttggcaatlaagcagacataggtgtcatgtctttttaccccttaaaaaaaattgttaa 1337
Db 1443 ACTTTGCCATTAAGCAGACATAGGGTTCAATGCTTTTAAACCCCTTTAAAAAAATTTGTA 1502
QY 1338 aagcttaagtaactacttttctcttgaatttgcagatttgacagcactcagaagcaat 1397
Db 1503 AAGTCTAGTACCTACTTTTCTTTGATTTTGACGCTTTGACTAGCCATCTCAAGCAACT 1562

QY	1398	ttcgacgcttgactagccatctcaagcaagtttaatacaagatcctacgctgacat	1457
QY	1458	tggaatcctactcaacaaaaggaagggtggtcgaagtaacattaaagattctgctccaa	1517
Db	1563	tttgcagcttttgacttaccatcttcacccaagttttaaaccaaaagatcatctcacgctgattat	1622
QY	1518	ttttcaataaattctctctctcctctcttaaaaaaaaaaaaaaaaaaaaaa	1563
Db	1683	ttttcaataaattctctctctcctctcttataaaaaaAAAAAAAAAAAAA	1728
RESULT	3		
LOCUS	116100	1749 bp	DNA
DEFINITION	Sequence 2 from patent US 5474901.		linear
ACCESSION	116100		
VERSION	116100.1	GI:1251008	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1749)		
TITLE	Dayna, D.T. and Eaton, D.L.		
JOURNAL	Antibodies to human carboxypeptidase B and methods of use thereof		
FEATURES	Patent: US 5474901-A 2 12-DEC-1995;		
	Location/Qualifiers		
	1..1749		
	source	/organism="unknown"	
BASE COUNT	521 a	361 c	342 g
ORIGIN			525 t
Query Match	87.2%;	Score 1372;	DB 6;
Best Local Similarity	90.3%;	Pred. NO. 4.3e-307;	
Matches 1558;	Conservative	0;	Mismatches 5;
			Indels 163;
			Gaps 2;
QY	1	agaaaaatctcttggatgaaagcctttgagcctctgcagctctgtaccatgtctctc	60
Db	24	agaaaattctcttggatgaaagcctttgagcctctgcagctctgtaccatgtctctc	83
QY	61	tctgtgagcagcaltgcttcgctgcctcagagtcgagcgaagttcagctgctctcctagaa	120
Db	84	tctgtgagcagcagatgcttctgcggtttcagagtcgagcgaagtttactgctctcctagaa	143
QY	121	cctctaagcgaagttcaagttctacgaatcttactacaacatagatgltctctggc	180
Db	144	cctctaagcgaagttcaagttctacgaatcttactacaacatagatgltctctggc	203
QY	181	agccggtiaacagctgacaccttatgtgaaagaaacaaagtcattttttgtaaagcat	240
Db	204	agccggtiaacagctgacaccttatgtgaaagaaacaaagtcattttttgtaaagcat	263
QY	241	ctgaatgacacatgltgaagcccatcttaaatgtgagcgaatccatgcagtgctctgc	300
Db	264	ctgaatgacacatgltgaagcccatcttaaatgtgagcgaatccatgcagtgctctgc	323
QY	301	tggcagaagctggaaagatcttatcaacagcagattccacagcaacagtcaccccgag	360
Db	324	tggcagaagctggaaagatcttatcaacagcagattccacagcaacagtcaccccgag	383
QY	361	cctcgacatgacacatgaacagatgaactcaactcaaatgaaatctattcttgatagat	420
Db	384	cctcgacatgacacatgaacagatgaactcaactcaaatgaaatctattcttgatagat	443
QY	421	ttataactgaagagatcctgatatgcttacaacaaatccacatggatcctcatgtgaa	480
Db	444	ttataactgaagagatcctgatatgcttacaacaaatccacatggatcctcatgtgaa	503
QY	461	agtaaccactctatgctttaaagttctcgtgaaagaaacaaacgcaaaatgacatat	540

D	504	AGTACCACTCTATGTGTTTTAAAGGTTTGTGGAAAAAGAACAAACGCAAAAATAGCCATAT	563
Q	541	ggattgactgtgnaatcgaatgcagagaatggaatctctcgtcttcgtcttggttca	600
D	564	GGATTGACTGTGGATTCATGCCAGAGAATGATGATCTCCGCTTTCTGTTGTGTTCA	623
Q	601	taggcaat-----	608
D	624	TAGCCATTTAACTCAATTCTATGGGATTAATAGGCCAATATACCAATCTCTGAGGCTTG	683
Q	609	-----	609
D	684	TGGATTTCTATGTATGCCGGTGTTATGTGAGCGTTATGACTCATCGAAAAAGA	743
Q	610	atcgaaatgtggagaagaacggtctcttctctatgcgaaacatcatgtcatgagaagacc	669
D	744	ATTCGATGTGGGAAAGAACCGTTCCTTTCTATGGGAACAATTCATTTGCATCGGAACGACC	803
Q	670	tgaatagaacttgcctcacaacatgtgtgagaagtgatccgaatctcctaagt	729
D	804	TGAAATGGAACTTGTCTTCCAAACACGTGTGTAGGAAGGTGCATCCAGTCTCTAGCT	863
Q	730	cggaacctactgtggaacttaacctgaatgaagaccagaagtgaagcaagtgtgctagt	789
D	864	CGGAACCTACTGTGTGACTTTATCTGTAGTCAGAACCAAGATGAAGGCGAGTGGCTAGTT	923
Q	790	tcttgaagaagaatacgaacagattaaagatatcatcagcatgcatatctccagc	849
D	924	TCTTGAAGAAATATCACACGATTTAAGCATATCATCATCAGCATCTCATCTCCACG	983
Q	850	atatagtttccatctctatacagaagtaaaagcaagaaccaatgagaactgtcc	909
D	984	ATATAGTGTTCCTCAATTTCTATACGAGATTAATAACCAAGCATGAGGAACGTCTTC	1044
Q	910	tgtatgcgaaggaagcagttgtgtctattgacaaaactagttaaaataccaggtataac	969
D	1044	TAGTACCAAGTAAAGCAATGTGTCTATGTAGAAAACCTACTAAATATACAGGTATAC	1103
Q	970	atgacacatggtctcagaacaccttaacatcctcgtgaggtgggagcgaatgtgactatg	1023
D	1104	ATGGCCATGGCTCAGAAACCTTATACGTAGTCTGTGGAGACATTTGGATCTATAG	1166
Q	1030	atttggacataaatatctgt-----	1056
D	1164	ATTTGGGATCAAAATATTCGTTTCAATTTGAACCTCGAGATACGGGCACATACGGATCT	1223
Q	1051	-----tcaacaacccaccgtatagaagcttttgcgcgtgtcctaata	1091
D	1224	TGCTCGCGGAGCGTTACATCAAAACCCACTGTAGAGAAGCTTTTGGCGTGTCTTA	1283
Q	1098	tagcttgcacatcatatagaatgtttaatgcccctgattatcaatctgtctcgat	1157
D	1284	TAGCTTGGCATGTCAATAGGAATGTTAAAGCCCGATTTTAATCATCTGCTCCGAT	1344
Q	1158	tttaattacatgattccaagaacgaacaaatcatgtatagaattatlttaagtttalc	1217
D	1344	TTTAAATTACGATTTCCAGCAAGCAACAAATCATGTATGATTAATTTTAAAGTTTATC	1403
Q	1218	cgtagtttgaataaagaattccatctctgttctgtctgcagagaacctaaatagtct	1277
D	1404	CGTAGTTTGTATTAAGATTTTCTATCTCTGTCTGTGAGAGAACTAATAGTGTCT	1463
Q	1278	actttgcataaaggcagaactaaggtcatgtcttttacccttaaaaaaaatgttaa	1337
D	1464	ACTTTGGCATTAAGGAGCACTAGAGTTCATGTCCTTTTAAACCTTTAAAAAAATTTGTA	1523
Q	1338	aagctatgtaacctactttctttaatcttgaagttgagatgaacatctcnaagcaact	1397
D	1524	AAGCTATGATTAACCTACTTTTCTTTGATTTTTCGACGTTTGACATAGCATCTCAAGCAACT	1583
Q	1398	ttcgaagtttgactagccatctcaagcaagttaatcaagaatcaatcnaagctatcat	1457
D	1584	TTTCACGTTTGACAGGCATCTCAAGCAAGTTTAATCAAAAGATCATCTACAGCTATGAT	1643

Qy 1518 ttccaataaattctctctccttaaaaaaaaaaaaaaaaaa 1563
 Db 1704 TTTCAATTAATTCTCTCTCTTAAAAA 1749
 RESULT 5
 BC007057 1715 bp mRNA linear PRI 12-JUL-2001
 LOCUS Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
 DEFINITION IMAGE:3934520, mRNA, complete cds.
 ACCESSION BC007057
 VERSION BC007057.1 GI:13937896
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1715)
 AUTHORS Strausberg, R.
 JOURNAL Direct Submission
 Submitted (30-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: CLONTECH
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mdc@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 16 Row: F Column: 21
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4503004.
 FEATURES
 source
 location/Qualifiers
 1..1715
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:12495 IMAGE:3934520"
 /tissue_type="Skeletal Muscle"
 /clone_1lb="NIH-MGC_81"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 25..1296
 /codon_start=-1
 /product="carboxypeptidase B2 (plasma)"
 /protein_id="AA07057.1"
 /db_xref="GI:13937897"
 /translation="MKLCSLAVLPVIVLCEQHVFAFQSGOVLALPRTSROYVLON
 LPTTVEIVLWQPTADLIVKQVDFVNSDVNKAHLNVGICSVLLAIVEDLI
 QOOISNDVSPRASASIDYEQHSLNEYSIEFTTBHPLMKRHIGSFEEYPIYV
 LKYSGEQAAKNIAWIDCGIHAREWISPAFLFIGHTQFVILIGQYTLALNPFY
 VMPVAVDGYDYSWKRNARSRKRSFYANNHCIGTDLNRRFAKNNCEBASSSSE
 TYCGILPESPESYKAVASTLRNINOKATISNHSYSHIVFYSYTRSKSDHEELS
 LVASEAVRALEKTSKNTRYTHGSEFLYLAPEGGDWYIDLGIKYSFTLELDGTGY
 GFLPRLRYKTPCREAFANVSKIAMVIRNV"
 BASE COUNT 518 a 348 c 341 g 508 t
 ORIGIN
 Query Match 83.2%; Score 1309; DB 9; Length 1715;
 Best Local Similarity 89.4%; Pred. No. 1.6e-292;
 Matches 1512; Conservative 0; Mismatches 15; Indels 164; Gaps 3;

Qy 1 agaaattgctgttggaatgaagcttgcagccttgacgtctctgtaccattgttctc 60
 Db 8 AGAAATGCTGTTGGATGAAGCTTTCAGCTTGCGCTCTGTATCCATGTGTTCTCT 67
 Qy 61 tctgtgaagcaagatgctcttcgcttcagagtgagcaagctctagctcctctcagaa 120
 Db 68 TCTGTGACAGCATGTCTTCGGGTTTCAGAGTGGCCAGTTCTAGCTCTTCTCTAGAA 127
 Qy 121 cctcgaagcaagttcaaatctctacagaatctctactacaacatgagattgtctctgagc 180
 Db 128 CCTTAGGCAAGTTCACAGTCTTACAGAACTTACTACAACTATAGAGATGTTCTCTGGCC 187
 Qy 181 agccgtaacagctgacattatgtgaaagaaaacaaagctctctttttttaaagcat 240
 Db 188 AGCCGGTAAAGCTGACCTTTATTGTGAAGAAACAAAGTCAATTTTGTAAATGATG 247
 Qy 241 ctgattgcgaatgtgaaagcccaattaaatgtgagcggaattcactgacagtgctctgc 300
 Db 248 CTGATGTCGACAAATGTGAAGCCCATTTAAATGTAGCGGAATTCATGACGTGCTTGC 307
 Qy 301 tggcagagtggaagatctttaaagcagagatttccaagacacagtcagcccgag 360
 Db 308 TGGCAGAGTGAAGATCTTATTCACAGCAGAACTTCCAAAGACACAGTCAGCCCGAG 367
 Qy 361 cctccgacatgctactatgaacagatcactcaactaaatgaactatctctgtgataag 420
 Db 368 CCTCCGACATGCTACTATGTAACAGATACATCAGTAATGAATTTATTTGGATTAAT 427
 Qy 421 ttataactgagagagcactcctgatactgcttaacaaatccacatgagtcctcattgaga 480
 Db 428 TTATTAACGAGAGGACATCTGATATGCTTACAAAATCCACATGTGATCCCATTTAGA 487
 Qy 481 agtaccacatagatttaaggttcttgaaagaaacaaagcgaataaagccat 540
 Db 488 AGTACCCACTATATGTTTAAAGGTTTCTGGAAGAAAGCAAGCCAAATAATCCATAT 547
 Qy 541 ggaattgactgtgaaatccatgccagagaatgagatctctctgtcttctgtgtgttca 600
 Db 548 GGATTTGACTGTGAATCATGCATGCCAGAAATGATCTCGTCTTGTTGTGTTCA 607
 Qy 601 taggcact----- 608
 Db 608 TAGGCCATATATCAATTTATGAGTAATAGGCAATATACCAATCTCCTAGAGCTTG 667
 Qy 609 -----a 609
 Db 668 TCGATTTCTATGTTATGCCGCTGTTATGATGATGATGATGATGATGATGATGATGAT 727
 Qy 610 atcgaatgtggaagaagaacccgtctctctctatgcaacaaatcattgacatggaagacc 669
 Db 728 ATCGAATGTGGAAGAAAGACCGTCTTCTATGCGAACAATATGATGATGATGATGATG 787
 Qy 670 tgaatgaacactgtctcctcaaacacactggtgtgaggaagtgatcgaatctctatgct 729
 Db 788 TGAATGGAACCTTGTCTTCCAAACACTGGTGTGAGGAAGTGCATCCAGTCTCATGCT 847
 Qy 730 cggaaacctactgtgacttatacctctagtcagaaacagagaagtgaaagcagtgactgt 789
 Db 848 CGGAACCTTACTGTGACTTTATCTGATGATGATGATGATGATGATGATGATGATGATG 907
 Qy 790 tcttgaaagaataatacaacagacttaagacatcagatcagatgacttactctcagc 849
 Db 908 TCTTGAAGAAATATATCAACAGATTTAAACATACATACATGATGATGATGATGATG 967
 Qy 850 atatagtgctcaattctctatacaggaagtaaaagcaagccatgtggaactgtctc 909
 Db 968 ATATAGTGTCTTCCATATTTCTTATACAGAAATGAAGCAACATGAGAACTGTCTTC 1027
 Qy 910 tagtaagcagtgaaagcagttcgtgctatgtgacaaactagtaaaatacaggtatcac 969
 Db 1028 TAGTAGCAGTGAAGCAGTTCGTGCTATGTGAGAAACTAGTAAATATACAGGTATACAC 1087

OY 970 atggccatggctcagaacattatagctcctcggaggtggagcattgattctatg 1029
|||||
Db 1088 ATGGCCATGGCTCAGAAACCTTATAGCTTGAGGTGGGACGATGGATCTATG 1147
OY 1030 atttggacatcaatcatcgt----- 1050
|||||
Db 1148 ATTTGGGATCAAAATTTGCTTTACATTTGAACTTCGAGATCGGCGACATCGATCT 1207
OY 1051 -----ttacatcaaacccacctgtagaagaactttgcgcgtctcttaaa 1097
|||||
Db 1208 TGGTCGGAGCGTTACATCAACACCACCTGTAGAGAACTTTTGGCGCTGCTCTAAAA 1267
OY 1098 tagcttgacatgcatcattgggaatgtttaatgccctgatttatactcttgcctat 1157
Db 1268 TAGCTTGGCATGTCATTTAGAAATGTTAAVGGCCCTGATTTTATCATTTGCTTCGGTAT 1327
OY 1158 ttaatttcctgattcccaagaacaaatcattgtatcagaattattttaagtttacc 1217
|||||
Db 1328 TTTAATTATCTATTCACGACCAACCAATCATTTGATACGATTAATTTAAGTTTATC 1387
OY 1218 cgtagtttgaataaagaatttccctatcctcgtgttcgtcagagaaacctaaatgct 1277
Db 1388 CGTATTTTGTGATTAAGATTTTCTATTCCTTGGTTCTGTACAGAACTTAATTAAGTCT 1447
OY 1278 acttggcattaaagcagactaggtgtcatgtcttttacccttataaaaaaatgttaa 1337
Db 1448 ACTTTTGCATTAAGGACAGCTVAGGGTTCATGCTTTTACCCCTTT-AAAAAAATTTGTA 1506
OY 1338 aagctagttacctaactttcttcttgatttcgaagtttgaacctcctaagaact 1397
Db 1507 AAGTATAGTACCTACTTTTCTTTGATTTTGACGTTTGATAGCACTCAAGCAACT 1566
OY 1398 ttgcagcttgcactagccatctcaagcaagttaactcaagaatcactcagcgtgcat 1457
Db 1567 TTGCAGCTTTGACCTACCTCAAGCAAGTTTGATGATGATCATCTCAGCGTGTAT 1626
OY 1458 tggatccctactcaaaaaaagaagggtgtgtcagaagtaacttaagaattctgctcaaa 1517
Db 1627 TGGATCTTACTCAACAAAAAGGAGGTGTGAGAGTACATTTAAAGATTTTGTCTCCATA 1686
OY 1518 ttccaataa 1528
|||||
Db 1687 AAAAAAAAAA 1697
RESULT 6
AR086324 1272 bp DNA linear PAT 07-SEP-2000
LOCUS AR086324
DEFINITION Sequence 1 from patent US 5985562.
ACCESSION AR086324
VERSION AR086324.1 GI:10013090
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Morser,M,John and Nagaashima,M.
TITLE Method of detecting thrombotic disease risk associated with plasma
carboxypeptidase B polymorphisms
JOURNAL Patent: US 5985562-A 1 16-NOV-1999;
FEATURES
source location/Qualifiers
1..1272
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN
Query Match 58.4%; Score 918; DB 6; Length 1272;
Best Local Similarity 86.8%; Pred. No. 4.7e-202;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;
OY 18 atgaagcttgagccttcagctccttgaccattgtctctctgtgtagaagcattgc 77
|||||

Db 1 ATGAACCTTTCGACGCTTGACGCTTGTATCCCATTTGTTCTCTCTGTACAGCATGTC 60
OY 78 ttgcgcttcacagatgagccaaagttcagctgtctcttccttagaacccttagcagaattcaa 137
|||||
Db 61 TTGGCGCTTTCAGAGTGGCCCAAGTTTACGCGCTTTCCTATGAAGACCTTAGGCAAGTTCAA 120
OY 138 gtctcagaatcttaactacaacatatgagatgtctctctgagccggttaacagctgac 197
|||||
Db 121 GTTCTACAAATCTTCTCAACATATGAGATTTGTTCTGTGGAGCGGTAACAGCTGAC 180
OY 198 cttaattgtaagaaaaaagaatcccaatttttgttaaatgcatctgatttgacaattg 257
|||||
Db 181 CTTATTGTGAAGAAAAAACAAGTCATTTTGTAAATGCACTGATGTGCACAATGTG 240
OY 258 aaagccatttaaatgtgacggaattccaatgacagatgctcgtcggcagaagtgagaat 317
|||||
Db 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTGTGGCAGACGTAAGAT 300
OY 318 ctatctcaacagacagatctccaagacacagltcagcccgagcctcgcatctactat 377
|||||
Db 301 CTTATTCAACAGCAGATTTTCCAAAGACAGTCAGGCCGCCGACCTCCCATCTGATAT 360
OY 378 gaacagtatcactcactaaatgaaatctatcttgatagaaatttaactgagagcat 437
|||||
Db 361 GAAAGATATCACTCACTAATGAAATCTATCTTGGATGAAATTTATTACTGAGAGCAT 420
OY 438 cctgatatgcttcaaaaaatccacattggaatccctcatcttgagaagtaccactatgtt 497
Db 421 CCGATATGCTTACAAAATTCACATTTGATTCCTCATTTTGAGAAATGCCACCTATGTT 480
OY 498 ttaaaagttcttggaagaagaacaacagccaaaatgcatatgattgactgtgaaac 557
|||||
Db 481 TTAAGATTTCTGGAAGAAACAAACAAACAGCCAAAAGCCATATGATGACTGTGGAATC 540
OY 558 catgccaagaaatgagatctcctcgtctctctgttctgttctgattgataagccat----- 608
Db 541 CATGCCAGAAATGATGATCTCTCTGCTTCTGTGTGTTGATAGCCCATATTAACTCAA 600
OY 609 ----- 608
Db 601 TTCTATGGGTAATAGGCAATATACCAATCTCCTAGGCTTGGAATTTCTATGTTATG 660
OY 609 -----aatcgaatgtgagaag 626
Db 661 CCGGTGTTAATGTGACGGTWTATGACTACTCATGGAAGAAAGATGATGAGAGAAAG 720
OY 627 aacctgtcttctatgacgaacaaatcattgcatgagacagacagcattgacaaattgctc 686
|||||
Db 721 AACGTTCTTCTATGCGAACAATCATTTGCATCGAACAAGACCTGAATAGGAATCTTGCT 780
OY 687 tccaacactggtgtgaggaaggtgcatcgaattcctcatgtctcggaaaacctactgtga 746
Db 781 TCCAAACTGCTGTAGGAGAGGTGCATCAGTTCTCATGCTCGGAAACCTACTGTGGA 840
OY 747 cttaatcctgagtcagaacagaagtgaagcagtggtcagtttcttgaagaagaatlc 806
|||||
Db 841 CTTTATCTGAGTCAGAACACAGAGGAGGCTGAGTTGTTCTTGAGAAAGAAATATTC 900
OY 807 aaccaattaaagcattacatcagcatgcatctatctatctccagatatagtgttccatct 866
Db 901 AACCAATTAAGCATATCATCATGATCATCTCATCTCCAGATATAGTGTTCATATAT 960
OY 867 tccatacagaagttaaaagcaagaacatgagaactgtctctatgtagccagtgaaga 926
|||||
Db 961 TCTATATACAGAAAGTAAAGCAAGACATGACAGAACTGTCTATAGCACTGAGAGCA 1020
OY 927 gtctgtctattgacaacaaactagtaaaaataccaggtatacacatggtcgtcagaag 986
|||||
Db 1021 GTTGTGCTATTGAGAAACTACTATAAATACAGGTATATACATGCGCATGTGCTAGAA 1080
OY 987 acctatagctcctctgaggtgtggagcagattgatactatgattgtgacaaatct 1046
|||||
Db 1081 ACCTTATACCTAGCTCTGTGAGGTGGGACGATTGGAATCATGATTTGGGACATCAATAT 1140

QY	1047	tcgt-----	-----	ttcc	1054
Db	1141	ttcgtttacattgaaacttcgagatagggacacatagatcttctgctggcggagcctttac			
QY	1055	atcaaacaccacgtctagagaagcttgcgcgtctctcaaaatagcttggcatgtcat			
Db	1201	atcaaacaccacgtctagagaagccttttggcgcgtctcttcttcttcttcttcttctt			
QY	1115	aggaaatgcttaa	1126		
Db	1261	aggaaatgctttaa	1272		
RESULT	7				
LOCUS	AB021968	1421 bp	mRNA	linear	ROD 27-JUL-2000
DEFINITION	Mus musculus mRNA for carboxypeptidase R, complete cds.				
VERSION	AB021968.1 GI:9558447				
KEYWORDS	carboxypeptidase R.				
SOURCE	Mus musculus cDNA to mRNA.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (sites)				
JOURNAL	Sato, T., Miwa, T., Akatsu, H., Matsukawa, N., Obata, K., Okada, N.,				
MEDLINE	Campbell, N. and Okada, H.				
REFERENCE	Pro-carboxypeptidase R is an acute phase protein in the mouse,				
AUTHORS	whereas carboxypeptidase R is not				
TITLE	J. Immunol. 165 (2), 1053-1058 (2000)				
JOURNAL	20341711				
FEATURES	2 (bases 1 to 1421)				
SOURCE	Sato, T.				
Location/Qualifiers	Direct Submission				
	Submitted (28-DEC-1998) Tomoo Sato, Fukushima Hospital, Chojiu				
	Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi				
	441-8124, Japan (E-mail: tomoctcp-1p.or.jp, Tel:81-532-46-7511,				
	Fax:81-532-46-4899)				
1..1421					
/organism="Mus musculus"					
/db_xref="taxon:10090"					
17..1285					
/codon_start=1					
/product="carboxypeptidase R"					
/protein_id="BAB01402.1"					
/db_xref="GI:9558448"					
/translation="MKLHGLIIVAILYEQHGFAFSGQVLSALPRTSROVOLLQNI					
TTTTTEVYLMQPVTAETFEIEKKKEVHFVNASVDVSKAHLNVRIPFVNLNNVDELTE					
QQTNDNVSPRASAYEYQSLNEISYMIWTEQHPDMQKLYISSPEKPYLVVL					
KVSKEDGRINKAIVIDGIIHREWISPAFCIMFYTGQFGEKNITRIILRHDPYIT					
MPVNVQGVDTYTKKRNMMRNKRBAHNRRVGVGIDNRPNASKMKSSSSDSETT					
YCGILPSEPEPVKAVADPLRNNDIHKAVISMHSYSQLPPYSYNSKSKSDHEEL					
VASAVSAVIESINKNTRTHGSGESLIVAGGSDMDIYDGLVSYPTIELRDGRG					
FLLEPERYIKPCALAAIAISKIYVHWYINRT"					
1421					
/note="11 a nucleotides"					
BASE COUNT	418 a	320 c	305 g	378 t	
ORIGIN					
Query Match	41.7%;	Score 656;	DB 10;	Length 1421;	
Best Local Similarity	72.2%;	Pred. No. 1.9e-141;			
Matches 1009;	Conservative	0;	Mismatches 215;	Indels 174;	Gaps 5;
QY	4	aaatgctgttggaagcttgacagcttgacagcttcgtctgtacaccatgctctctct	63		
Db	3	aagtcactgttggaagcttcacgttcacgttcggaatcctggagaccatcactc	60		
QY	64	gtgaagcagcatgcttcgcttcagagtgccaagtcttagctgcttctctagaacct	123		
Db	61	-tgacacacatggccttcctccctttacagactggccacagctttttatgcctttcc	119		

QY	124	ctaaaggaaagttcaagttcttacagaactcttactaacatactgaagattgtctctcggcagc	183
Db	120	CCAGGCAATGTTCAACTACTTCTGAAATCTTTACTACAGGTATGAGGTCTCTCTGGCAGC	179
QY	184	cggtaacgctgactctattgtggaagaaaaaagaatccatttttctgtaatgcaatctg	243
Db	180	CAGTGCACACTGAAATTCATCGAGAAAGAAAAGAAAGTCACACTTTTGTGAAAGCGCTCTG	239
QY	244	atgtcgacaacatgtgaagcccatttaaagtgtgacgggaattccatgagtgctctgtg	303
Db	240	ATGTGACAGCTGTCAAAACCGCATTTAAATGTGACACAAATTCCTTTAACTGTTCTATGA	299
QY	304	cagacgttggaagatcttaataacagcagaattccaaacgacagtcgaagcccgagcct	363
Db	300	ACAAGGTGAGGAGCACTAATTGAACAGCAGACTTTCATATGACAGGTGACGCCGGCGCT	359
QY	364	ccgacgtgactatgaagaagatatactatccatccaaatgaaatctattcttggatgaatta	423
Db	360	CCGCTTCACTATGATGAGAGATACGTACTCGCTAAATGGAATCTTTCTCGATGGAAGTGA	419
QY	424	taaacgagaagcattccctgatatgtcttccaaanaatcccatctggatccctcaattggaaagt	483
Db	420	TAACTGAACAGAGATCCGTACATGCTCCAGAAAATCTACATCGATATCATCTTTGACAAGT	479
QY	484	accacactatgattttaaaggttctctggaabaagacaacagcnaaaatgycataatga	543
Db	480	ACCCACTTATGTTGTTTAAAGGTCTCAGGAAAGGAAAGAAACAAATCAAAAATATGCGATCGGA	539
QY	544	ttgactctggaatcccatgacagagaatggatctctctgtcttctctgttggttcatg	603
Db	540	TGCACTGTGGAAATCCATGCCAGAAATGATTTTCACCTGCTTCTGTGTGTGTTGATATAG	599
QY	604	gc-----	605
Db	600	GCTAGCTGACACAATTCATGGAAGAAATAATCTGTATACAGACTTCTGAGCAGAGTGG	659
QY	606	-----cataatc	612
Db	660	ATTTCATCATATGCGCGTGATGAGAGTGGATGCTATGACTACACGTGAGAAAAGAAATC	719
QY	613	gaatgtggaagaagaacccgtctcttctcatgtcgacaacaatcatgtcatcggaacagaccta	672
Db	720	GAATGTGAGGAGAAACCGCTCTCTCTACAAAGAAACAACCGCTCGTGGGCACAGACTGA	779
QY	673	atagcaacttgtctccaacaactggtgtggaagaagtgcatacagttccctcatgtcgg	732
Db	780	ACAGCAATCTGCTTCCAAACACACGTGTTGAGAAAGGTGCGTCAATGTCCTCGTGCTG	839
QY	733	aaacctactgtggacttatcttcctgagtcacaacaacgaagtgaagagagtgctagtcttc	792
Db	840	AAACCTACTGTGGACTTATTCCTGATGCTGACGACAGGGGAGGAGCGAGTGGCTGACTTCT	899
QY	793	tgaagaagaataataacacagattaaagacatactagcatgcatctcatcactccagata	852
Db	900	TGAGAGAAATATGACACACTTAAAGCTTACATCGATGACATCATCTCCCAACAA	959
QY	853	tagtgttccatattcctatatacacgaagttaaagaagaacatggaagaaatgtctctag	912
Db	960	TACTGTTTCCCTATTCCTATTAACAGAAAGCAAAAGCAAGACACAGAAATGTTCTTAG	1019
QY	913	tagcagatggaagcaggttcggtgcatatgaacaaactagttaaaatcccgattataacatg	972
Db	1020	TGGCCACGAGAAAGCATTCGTGCGCATTTGAAGTATTTAATTAANAACACAGATACACACG	1079
QY	973	gcataggctcagaacacttaactagctctcgtgaggttgaggagatgtagctatagtt	1032
Db	1080	GCACTGGCTGAGAAAGTTTATATCTAGCTCTCTGAGGTTCTGACATTTGAGATATGATT	1139
QY	1033	tgaggcatccaatatgtgt-----	1051
Db	1140	TGGGCATCAATAATTCCTTTACAATTGAGCTCCGAGATACAGCAGAGATACGATTTCTTC	1199

QY 1052 -----tacatcaaacccacctgtatgaagacttttgcgcgtctctctaaatag 1100
Db 1200 TGCGTAGAGATACATCAACACCTGCTGCGAGAGCTTTGGCCGACCTCTTAATAATG 1259
QY 1101 cttagcagctcaatgaagatgttaatgccccctgatttcaatcctgcgtctatctt 1160
Db 1260 TTTGGCAGTGTCAACGAGCACTTAATGCCC-----TAACTCCGCTCTCATTTATTTT 1313
QY 1161 aattactgattccagagaccaaataatcattgtaacagat--taatttaagttatcgc 1218
Db 1314 ATTATATGATTTGACGACAACTTAATGCTGTTCATTAAGTCTTAAGTGAATCAGTTCC 1373
QY 1219 gtatgttgcataaagat 1236
Db 1374 TTGGTTTGTTCAGAAAT 1391

RESULT 8
AF164524 1490 bp mRNA linear ROD 05-APR-2000
LOCUS AF164524
DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (Taf1)
ACCESSION AF164524.1 GI:7416966
VERSION AF164524.1
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1490)
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
AUTHORS Marx,P.F., Wagenaar,G.T.M., Reijerkerk,A., Tiekstra,M.J., van Rossum,A.G.S.H., Gebbink,M.F.G.B. and Meijers,J.C.M.
TITLE Characterization of mouse thrombin-activatable fibrinolysis inhibitor
JOURNAL Thromb. Haemost. 83 (2), 297-303 (2000)
MEDLINE 20201996
REFERENCE 2 (bases 1 to 1490)
Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
AUTHORS Direct Submission
JOURNAL Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100, Utrecht 3584 CX, The Netherlands
FEATURES
source location/Qualifiers
1..1490
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:477035"
1..1490
/gene="Taf1"
84..1352
/gene="Taf1"
/function="carboxypeptidase"
/product="thrombin-activatable fibrinolysis inhibitor"
/protein_id="AA162385.1"
/db_xref="GI:7416967"
/translation="MKLHGILVAIIILYEQHGAFSGQVLSALPRTSROYVLQNL
TTTEVYLMQVTAIEFEIKKKEVFNASVDVSKALNRSIPFNLMNVEDLIE
QOTFNDSPPASASAYEIOYSLNEIYFVIAEMIEVIOHDMLOKIVIGSSFEYPLVL
KVSGEORIKNAIWDGIIHAREMISPAFCMEFIEGYVOFGKELVYRLIHLVDYFI
MPVNVQSDYDTWKRMKMRKNSAHKRNRCVGDNLNPKSKHMLCEGASSSCSET
YCGLPSEPEPKAVADFLRNIDIKIYISMHSISQILFFYSINRSKDHDELST
VASEAVRAIESINNTRTTSGSSESLYLAPGSDWDWIDIGIKYSFTLELDTRNG
FLPERYKRPICAAALAIISKIWMHVINT"

BASE COUNT 438 a 337 c 322 g 393 t
ORIGIN

Query Match 41.7%; Score 656; DB 10; Length 1490;
Best Local Similarity 72.2%; Pred. No. 1.9e-141;
Matches 1009; Conservative 0; Mismatches 215; Indels 174; Gaps 5;

QY 4 aatgctgttgatgaagcttgacgctgcagctcttaccatgttctctct 63
||| ||||||||||||||| ||||||| ||||| ||||| |||||

Db 70 AACTCACTGTTGGAGATGAAGCTTCATGCGCTTGGAATCCTGGTAGCCATCATCTCTTA-- 127
QY 64 gtgagcagatgtcttcgcgttcacagagtgagccaaagttcagctcttcacagacct 123
Db 128 -TGAGCAGATGGCTTCGGCTTCAGAGTGGCCAGGTTTATCTGCTTCGAAGAACCT 186
QY 124 ctgagcagatctcaagttcacaagaatcttactacaacatatgatatgttctctgcagc 183
Db 187 CCAGGCAAGTTCAACTACTTCACAAATCTTACTACAAACGATGAGGCTGCTCTGACAC 246
QY 184 cgttaacatgacatctatgtgaagaaaaaaagttccatctttttgttaaatgacatc 243
Db 247 CAGTGACAGCTGATTCATTCGAGAGAAAAGGAAGTCCACTTTTGTGTAAGCCGTCTG 306
QY 244 atgtcacaatgtgaagccatcttaaatgtgagcgaatcccatgcagtgtctctgcg 303
Db 307 ATGTGACAGTGTCAAGGCAATTTAAATGTGAGAGAAATTCATTTAAGCTTCATATG 366
QY 304 cagacgtggaagatccttattcaacagcagattccaaacgacagtcagccccgcgct 363
Db 367 ACNACGTGAGAGACCTTAATTGAACACAGACTTTCATGACAGGTCAGCCCGCGCCT 426
QY 364 cggcactgactatgacagatcactcaactcaataatgaaatctatttgatagaattta 423
Db 427 CCGCTTCATATCTATGAGACAGATATCAGTCCGTAATGAATCTATTCCTGGATGAAGTCA 486
QY 424 taactgagagcactcctgatatgcttaccaaaataccacatlgatcctcatttgagaagt 483
Db 487 TAAGTGAACAGATCCTGACATGCTCCAGAAATCTACATCGGATCATCTATGAGAGAT 546
QY 484 acccactatgttttaaaagttctggaagaagaacaaacagcacaataatgcatatgga 543
Db 547 ACCCACTTATGTTTAAAGGCTCAGGAAGAAAGAAAGAAATCAAAAATGCCATCTGGA 606
QY 544 ttgactgtgaaacatcttcgcagaaagatgctcctctgcttctctgttggttaag 603
Db 607 TCAGCTGTGAATCCATGCGCAGAAAGGATTTACCTGCTTCTGTGTGTGTTCAATAG 666
QY 604 gc----- 605
Db 667 GCTNACGTGACAAATTCATGGAAGAAATACTGTATACACAGCTTCTGAGGACAGTGG 726
QY 606 -----cataatc 612
Db 727 ATTTTACATCATGCCGCTGATGAAGCTGATGCTATGACTACAGTGAAGAAAGATC 786
QY 613 gaatgtggaagaagaaccttcttctatgcaagaatacttgatcagcagaagacttga 672
Db 787 GAATGTGAGAGAAACCGCTGCTCACAGAAACACCGCTGCGTGGCAGACCTGGA 846
QY 673 atagcaacttgtctcacaacactgtgtgagaagaagtgatcagttcctcatgtcgg 722
Db 847 ACAGGAACCTCGCTTCCAAACATGAGTGTGAAGAAAGTGGTGAAGTCTCTCGTCTG 906
QY 733 aaactactgttgacttattcctgtgagtcagaccagaagtgaagcagtgagttct 732
Db 907 AAACCTACTGCTGACCTTAATCCTGAGCTGACGACAGAGTGGAAGGCGTGGCTACTTC 966
QY 793 ttgaggaagaatatacaacagagataaagcatatcagatgcatcttaactccagcata 852
Db 967 TGAAGAAGAAATATGACACCATTAAGCTTACATGACTATGACTCATCTCCACAAATA 1026
QY 853 taatgttccataatctctatacagaagtaagaagcagaagcctatggaacttctctag 912
Db 1027 TACTGTTTCCCTATTCCTATTAACAGAAAGCAAGAGCAAGAAATCTGCTCTAG 1086
QY 913 tagccagtgagcagatgtgtgtatgtgacaacaaactgtaaaataaccaggtatacacatg 972
Db 1087 TGCGCAGCGAGAACGATTCGCGCAATGGAATGAAATTAATAAAAAACACAGGTACACACAGC 1146
QY 973 gcatgtcctagaagaaccttactacagctccttggaagtgaggatgtgatatgatt 1032
Db 1147 GCAGTGGCTCAGAAAGTTTATATCTAGCTCTCGAGGTTCTGACGATTTGATCTATGATT 1206

QY 1033 tgggcatcaaatctcgt----- 1051
 |||||||
 Db 1207 TGGGCAATCAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAAGATACGATTTCTTGC 1266
 QY 1052 -----tacaacaacccactgttagaagaactttgcgcgtctcttaaatag 1100
 |||||||
 Db 1267 TGCCTGAGATACATCAATCAATGCTTGTGAGAGCTTGGCGCCATCTCTAAATAG 1326
 QY 1101 ctggcagtcagtcagtaagtgcttaagtcgcccagtttcaatcttcgcttcgatttt 1160
 |||||||
 Db 1327 TTGGCATGTCATCAGAACACCTTAATGCC-----TAACTCCGCTCTCATATTTTT 1380
 QY 1161 aattctagtcagtcagcaagaacacatctgtaacagat--tattttaagttatcc 1218
 |||||||
 Db 1381 ATTATATGATTTCAGCAACACTTAAGTTCATAGCTTCCTAAGTTGAATCAGTTTCC 1440
 QY 1219 gtagtttgcataaagaat 1236
 |||||||
 Db 1441 TTGGTTTGGTGAGAAAT 1458

RESULT 9
 AF186188 1430 bp mRNA linear ROD 01-OCT-1999
 LOCUS AF186188 Mus musculus carboxypeptidase U mRNA, complete cds.
 DEFINITION AF186188
 ACCESSION AF186188.1 GI:6003651
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 house mouse
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS He,Y.C. and Broze,G.
 TITLE Isolation and characterization of mouse liver carboxypeptidase B
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1430)
 AUTHORS He,Y.C. and Broze,G.
 TITLE Direct Submission
 JOURNAL Submitted (13-SEP-1999) Division of Hematology, Washington
 University School of Medicine, 216 S. Kingshighway, St. Louis, MO
 63110, USA

FEATURES
 source
 1..1430
 /location="Qualifiers
 /organism="Mus musculus"
 /strain="B6/CBAF1J"
 /db_xref="taxon:10090"
 20..1288
 /note="pro-carboxypeptidase-B; thrombin-activatable
 fibrinolysis inhibitor (TAFI); pCPU; pCPB; binds
 plasminogen"
 /product="carboxypeptidase U"
 /protein_id="AF00528.1"
 /db_xref="GI:6003652"
 /translation="MKLHGIGIIVAILLYEONHGFAROSQVLSALPRTSROVLONL
 TTYEVLMOPVAEIEFKKLVFVFNASDVSAHLNVSILPENVLMNNEDILE
 OOTFNDTVPRASAYEIOYHSLNLEYEVLTEDPDLKITYGSSPEKPLVLI
 KYSGKQRIKNAIWDICGTHARHISPAICWITGYTORHGENYTRLLRVDVYI
 MPVANYDGTDTMKRNRKRNKRSARKNRNCVTDLNKRNASKHKEKGSSSCSET
 YCGLYPESEPEVAVADFLRNIDHIAKVISMSYSQOILFPYSNRSKSKDHEEISL
 VSAEAVRAIGSFNKNTRYTHGSGESLILAPGCGSDMIYDLGIKYSTIELRDTGRYG
 FLPERYIKPTCEALAAISKIWMHVIKMT"

BASE COUNT 421 a 320 c 310 g 379 t

Query Match 41.6%; Score 654; DB 10; Length 1430;
 Best local Similarity 72.4%; Pred. No. 5.6e-141;
 Matches 989; Conservative 0; Mismatches 205; Indels 172; Gaps 4;

QY 1 agaaatgtcgttggatgtagaagcttgcagcccttgacccatgttctct 60
 |||||||
 Db 3 AGAAGTCACATGCTGGATGAACTCATGGCTTGAATCTCGTATCATCTCT 62
 QY 61 tctgtgagcagcagcttctgcgcttcagagtgtagaagcttgcgtctctcaga 120
 |||||||
 Db 63 A--TGAGCACATAGGCTCGCTTTCAGAGTGCCAGGTTTATCTGCTTCCAGAA 119
 QY 121 cctctagcagcttcaagcttctacagaactcttaacacacataagatgtctcggc 180
 |||||||
 Db 120 CCTCAGGCAAGTTCACACTTCTCAGAACTTACTACAGCTATGAGGTGCTTCTGCGC 179
 QY 181 agccggttaacagctgaccttatgtgaaagaaaaaagatcatttlttgtaatgat 240
 |||||||
 Db 180 AGCCAGTCAGCTGAATTCATGAGAAAGAAAGAAAGTCCACTTTTCTGTAATGCT 239
 QY 241 ctgagtgcagacatgtgaaagccatttaagtgtgagcgagatccatgacgtgtctgc 300
 |||||||
 Db 240 CTGATGTCAGACAGTGTCAAAAGCATTAAATGTGAGCAGAAATTCATTACGTTCTGA 299
 QY 301 tggcagacgttggagatcttatacaacagcagatctccaacagacagtcagcccgag 360
 |||||||
 Db 300 TGAACAACGTGAGAGACTAATTTGAACAGCAGACTTTCATATGACAGCGTCAAGCCCGCG 359
 QY 361 cctccgcatgtactatgaaacagatatacctaactaatgtaacttatcttgatagaat 420
 |||||||
 Db 360 CCTCCGCTTCACTATGATGAGCAGATCATCTGCTAAATGAATGAAATTCCTGGATAGAG 419
 QY 421 ttataactgagagcagctcgtatgcttacaacaaatccacatggtatcccatgaga 480
 |||||||
 Db 420 TCTTACTGAACACAGCTCTACATGCTCCAGAAATCTACATGAGATCATATTCCAGA 479
 QY 481 agtaccactatgtttaaaggcttcgtgaaagaaacaaacagccaaatgtccatat 540
 |||||||
 Db 480 AGTACCACCTTATGTTTAAAGTCTCAGAGAAAGAAAGAAATCAAAAATGCCATCT 539
 QY 541 ggaatgactgtggaatcagtcagcagagatgagatctctcgtcttctgctgtgttca 600
 |||||||
 Db 540 GGATCGACTGTGGAATCATGCCAGAGATGAAATGAAATTCACCTGTTGTGTGTTCA 599
 QY 601 taggc----- 605
 |||||||
 Db 600 TAGGCTACGTACACAAATTCATGTGGAAGAAATCTGTATACAGACTTGTGAGGACG 659
 QY 606 -----cata 609
 |||||||
 Db 660 TGAATTTCTACATCATGCGCGCTGATGAACGTGATGCTATGACTACACGTGAAAGAA 719
 QY 610 atcgaatgtggaagaagaacgcttcttctatgcgaacatcatctgacagagacc 669
 |||||||
 Db 720 ATCGAATGTGAGGAGAACCGCTCTGCTCAAGAAACAAACCGCTGCGTGCGGACAGAC 779
 QY 670 tgaatagaacttgcctccaacacatgtgtggaagagtgatccagctcccatgct 729
 |||||||
 Db 780 TGAACAGGAATTTGGCTTCCAAACACTGTGTGAGAAAGTGCCTCAAGTCTCTCTGCT 839
 QY 730 cggaaacctactgtgacatttaccctgagtcagaacccaagtgaaagcagtggtcgt 789
 |||||||
 Db 840 CTGAACCTACTGTGGAATTCCTGAGTCTGAGCCAGAGAGTGAAGCACTGCTGCTGACT 899
 QY 790 tcttgagaagaataataacacagatataagatcatcatcatcatcatcatcatccagc 849
 |||||||
 Db 900 TCTTGAGAGAAATATGACACCAATTTAAAGTTTCAATCATGATCATCATCATCTCCAAC 959
 QY 850 atatagtttccatattccatacagaaftaaagaagaacatagaaactgtctc 909
 |||||||
 Db 960 AATATCTGTTTCCCTATTTCTATTAACAGAACCAAAAGCAACGACACCAAGAACTGTCTC 1019
 QY 910 tagtagccagtgaagcagcttctgtctatgacaaactagtaaaaaataccagataaac 969
 |||||||
 Db 1020 TAGTGGCCAGGAGAGCAAGTGTGCAATTTGAAGTTTAAATAAACAACACAGATACACAC 1079
 QY 970 atggcatggtctcagaacactataactagctccttgagaggtcgaggagatgagatcatg 1029

Accession	Sequence	Position
Dh	 ACGGCAGTGGCTCACAAGATTATATATCTAGCTCCTGGAGTTCTGACGATTGGATCTATG	1139
Qy	atttggagatcaaatattcgtt-----	1051
Dh	1140 ATTTGGGCATCAAAATATTCGTTTACAAATTGAGCTCCGAGATACAGGCAAGATACGATTCT	1199
Qy	1052 -----tactcaaacccaccctgtagaaagcttttgcgcgtctcttaaa	1097
Dh	1200 TGTCTCCTCGAGAGATCAATCAAAACCCACTTTGGCAGAAAGCTTTGGCCGCCCACTCTTAAA	1259
Qy	1098 tagcttgacatgcatacttagaattgttaatgcccctgatttcatcatctgcttcgcat	1157
Dh	1260 TAGTTTGGCAGTGCATCAAGAACTATTATGCCC-----TAACCTCGCGTCCTATTATT	1313
Qy	1158 tttaatttactgattccagcaagaccaaatcatgltatcagattat	1203
Dh	1314 TTTATTATTAATGATTTTCAGCAACACTTCACTGTTTCATTTAGCTTCT	1359

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL
1 (sites)	Kato,T., Akatsu,H., Sato,T., Matsuo,S., Yamamoto,T., Campbell,W., Hotta,N., Okada,N. and Okada,H.	Molecular cloning and partial characterization of rat procarnboxypeptidase R and carboxypeptidase N	Microbiol. Immunol.	44 (8), 719-728 (2000)	20471387	2 (bases 1 to 1425)	Kato,T.	Direct Submission
							Submitted (10-MAY-2000)	Tomomi Kato, Chojiu Medical Institute, Chojiu

FEATURES	Location/Qualifiers
source	1. 1425 /organism="Batus nortoniensis"

CDS	100.1368
-----	----------

BASE COUNT	417 a	334 c	309 g	365 t
ORIGIN	BLUE EYE LINE CREEDMANVSNALMYLAND			

	Query Match	41.5%	Score 652.6;	DB 10;	Length 1425;
Best Local Similarity	72.8%;	Pred. No. 1.2e-140;			
Matches 981; Conservative	0;	Mismatches 194;	Indels 172;	Gaps 4;	
QY	4	aaatgcgtgtggaggaagcttcgacgcttgacgcttcctctgttacocatgtttctctct	63		

Db	86	AAGCACGTTGGATGGAAGCTTTATGCGCTTGAGACGCGTGAAGCCATCATCTCTA--	143
Oy	64	gtgaagcaatgcttctcggttccagaatgccaagtcttactgctcttctagaact	123
Db	144	-TGGAAGCATGGCTTGCTTCAGAGTGGCCATGTTCATCTGTCTCCGCAACT	202
Oy	124	ctagaagaattcaagttctcagaacttactacaatatgagattgtctctcagaac	183
Db	203	CCAGGCAAGTTCAACTTCTTCAGATCTCACTACAACTTACGAGGTGTCTCTGGCAC	262
Oy	184	cggtlaacagctgacacttatgtgaaagaaaaaagllcaatlttttgttaatgcatg	243
Db	263	CAGTACAGCTGAATTCATTTGAGAGAAAAAGAAAGCACTCTTTGTGAATGCGCTG	322
Oy	244	atgtcgcgaatgtgaaagccccattaaatgtgagcggaattcgaatgcatgtcttcg	303
Db	323	ATGTCAACAGTGTCAAGCCATTATTTAAATCGAGCAAGAAATTCATTAACTGATTA	382
Oy	304	cagacgtggaagatcttattccaacagcagaattccaacgcacacagtcagcccgagct	363
Db	383	ACAAGCGGAGGATCTAAATTCACAGCAGACGTCGAATGACACTGTACCCCGAGCCT	442
Oy	364	ccgcacgtactatgaaacaglatcaactcaataaagaatctatctcttgatagaatta	423
Db	443	CCCTCTATCTATGTAAGCAAGTATCACTCGCTTAATGAATCACTATTCCTGGATGAAGTTA	502
Oy	424	taactggaagagcatccctgatatcttcaaaaaatccacattgatatcttattgaagaat	483
Db	503	TAACTGAAACAGCACCCCTGACATGCTCCAGAAAATCTACATTTGGATCTCATATGA	562
Oy	484	accacacttatgttttaaaagttctctgaaagaagaacaaacagccaaaaatgycata	543
Db	563	ACCACATTATGTGTAAAGTCTCAGAGAAAGAACACAAAGTCAAAAATGCGCATATGA	622
Oy	544	ttgactgtgaaatcatcgacagagaatggatctctcgtcttcttggttgatag	603
Db	623	TGCACTGTGGAATCATGCGCAGAGAGTGAATTCACCAAGCTTCTCTGTTGGTTCATAG	682
Oy	604	gcat-----	608
Db	683	GCTATGTAAACCAATTCCTATGGGAAAAAATAATCATACACAGACTTCGAGGCAAGTGG	742
Oy	609	-----aatc	612
Db	743	ATTTCATATTATGCGAGTATGATGTGACGCGCTACGACTACAGTGGAAAAAATAATC	802
Oy	613	gaatgtgagaagaagaaccttcttctatctctcagacatactatgcatcggaagagactga	672
Db	803	GAAATGTGAGAAACACCGCTCTGTCCACATGAAACAACCGCTCTGTGGCACAAGACTTA	862
Oy	673	atagaacattgttccaaaacactggtgtgaggaaggtgcatccagtctctcatgtctgg	732
Db	863	ACAGGAAGCTTGCTTCCAAAACACTGGGTGAGAAAAGCGCATCAATTTCTCTCGCTGTG	922
Oy	733	aaactctctgagacttatctctgaatgcagaacacagaagtgaaagcagatggcctaattct	792
Db	923	AGACCTACTGTGGACTTATACCTCTAGTCTAGCCACAGAGTGAAGGCAATGGCTGACTTTC	982
Oy	793	tgaagaagaataataaacacagattaaagcatatacagcatgcatatccatccacata	852
Db	983	TGAGAGAAATATCAACACATTAAAGCTTACATAGTATGCACTCATCTCCACAGAA	1042
Oy	853	taagtgttcatatctctatacagaagtaaaagcaagcatgaggaactgagctctctag	912
Db	1043	TACTGTTTCCCTATTCCTACAAAGAAAGCAAAAGCAAGCAACGAGAAATGTCTCTAG	1102
Oy	913	taagcagtgaagagactctgctgctatgacaaaaactgttaanaataccaggtatacacatg	972
Db	1103	TGGCCAGGAAGAGCTTCGTGCCATTTAAGTATTAATTAATAAACACCAAGTACACACATG	1162
Oy	973	gcacatgctcagaacattacactagctctcggagagtgggagcattgatcatgatatt	1032
Db	1163	GCATGTGGCTCGAAGATTTATATCTAGCTCTGAGAGTGTGTGATGATGGATCTATGATT	1222


```

Oy 1033 tgggcataaattatcgttt----- 1053
Db 1223 TGGGCATCAAAATATTCGTTTACGATTGACATTCGGGATACAGGAGATACGGGTTCTTGC 1282
Oy 1054 -----catcaaacccacactctagaaagctctttgcgcgtctctctaaatatag 1100
Db 1283 TGCGTGAGAGATATCATCAAAACCACTTCGCGAGAAGCTTTGGCCGCACTGTAATAAATAG 1342
Oy 1101 ctggcattgctattcgaatgtttatgcccctgatttalcattcgtctccgtatttt 1160
Db 1343 CTGGCAGTCTCATCGGAACAAGTTAAACACC-----TTTCCTCTGCTCATACTTCTTT 1396
Oy 1161 aattactgattccagcaagaccaaat 1187
Db 1397 ATTTATGATTTCAGCAACACTAAAT 1423

RESULT 11
LOCUS AL157758/c
DEFINITION Human DNA sequence from clone Rp11-106H11 on chromosome 13,
ACCESSION AL157758
VERSION AL157758
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 58097)
AUTHORS Griffiths, C.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humgen@esanger.ac.uk
COMMENT Requesters: clonerequests@sanger.ac.uk
On Jan 21, 2001 this sequence version replaced g1:12189591.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
Chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
Rp11-106H11 is from the library RPC1-11.1 constructed by the group
of Pleter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pACe3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-106H11 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone Rp11-139H14 is at 57998 in this
sequence. The true right end of clone Rp11-71J12 is at 100 in this
sequence.

FEATURES
Source
1..58097
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="Rp11-106H11"
Location/Qualifiers

```

misc_feature	/clone_11b="RPCI-11.1"	complement(1..95)	
repeat_region	/note="match: GSS: Em:AQ266238"	624..851	
repeat_region	/note="L1P48 repeat: matches 5955..6162 of consensus"	1331..1611	
misc_feature	/note="L1MC/D repeat: matches 5541..5822 of consensus"	complement(1912..2370)	
repeat_region	/note="match: GSS: Em:AQ335999"	2288..3195	
repeat_region	/note="L1M2A repeat: matches 5203..6161 of consensus"	3676..4108	
repeat_region	/note="Char11eb repeat: matches 2..463 of consensus"	4453..4871	
misc_feature	/note="12 repeat: matches 1151..1600 of consensus"	complement(4551..5058)	
repeat_region	/note="match: GSS: Em:AQ662290"	5674..5855	
repeat_region	/note="MRS5A repeat: matches 1..189 of consensus"	6139..6317	
repeat_region	/note="AluJo repeat: matches 16..193 of consensus"	6318..6416	
repeat_region	/note="AluDb repeat: matches 194..293 of consensus"	9114..9392	
repeat_region	/note="MRA4C repeat: matches 23..333 of consensus"	10838..11072	
repeat_region	/note="HAL1 repeat: matches 571..819 of consensus"	11095..11216	
repeat_region	/note="L1MCC repeat: matches 1471..1593 of consensus"	11771..12046	
repeat_region	/note="AluJo repeat: matches 12..284 of consensus"	12651..12772	
repeat_region	/note="41 copies 2 mer aa 638 conserved"	12920..13482	
repeat_region	/note="L1M4 repeat: matches 2645..3199 of consensus"	13549..13850	
repeat_region	/note="AluDb repeat: matches 3..305 of consensus"	13884..14086	
repeat_region	/note="L1M4 repeat: matches 3359..3566 of consensus"	14322..15872	
repeat_region	/note="L1MCC repeat: matches 2383..3622 of consensus"	15873..16175	
repeat_region	/note="AluJ repeat: matches 1..303 of consensus"	16176..16335	
repeat_region	/note="L1MCC repeat: matches 3622..3777 of consensus"	16316..16419	
repeat_region	/note="L1M49 repeat: matches 6158..6264 of consensus"	16425..18501	
repeat_region	/note="L1MCC/D repeat: matches 3729..5776 of consensus"	18517..18700	
repeat_region	/note="L1MCI repeat: matches 6134..6318 of consensus"	18790..18835	
repeat_region	/note="23 copies 2 mer aa 788 conserved"	19057..19377	
repeat_region	/note="L1ME repeat: matches 5607..5930 of consensus"	20066..20261	
repeat_region	/note="L1MCC/D repeat: matches 5280..5504 of consensus"	20263..21196	
repeat_region	/note="L1MEL repeat: matches 5133..6163 of consensus"	21376..21509	
repeat_region	/note="L1M23 repeat: matches 6602..6740 of consensus"	21570..21822	
repeat_region	/note="MT1H repeat: matches 227..490 of consensus"	21578..22061	
misc_feature	/note="match: GSS: Em:AQ697967"	22450..22616	
repeat_region	/note="AluDb repeat: matches 137..303 of consensus"	23172..23466	
repeat_region	/note="AluX repeat: matches 1..305 of consensus"	23539..23673	
repeat_region	/note="L1M24 repeat: matches 6043..6182 of consensus"	24142..24404	
repeat_region	/note="MER33 repeat: matches 53..324 of consensus"		


```

repeat_region 24437..24464
/note="MER5B repeat: matches 1..38 of consensus"
24493..24511
/note="L2 repeat: matches 2606..2659 of consensus"
repeat_region 25199..25476
/note="AluSg repeat: matches 11..287 of consensus"
26044..26421
/note="MER63B repeat: matches 1..435 of consensus"
26438..26592
/note="MER5A repeat: matches 2..149 of consensus"
26821..28021
/note="L1MB8 repeat: matches 4920..6157 of consensus"
28733..28774
/note="21 copies 2 mer aa 76% conserved"
/note="complement(30583..31098)"
/note="match: GSS: Em:A0507270"
31066..31450
/note="match: GSS: Em:A0094491"
32825..33131
/note="AluY repeat: matches 1..297 of consensus"
34745..35053
/note="AluY repeat: matches 2..306 of consensus"
35297..35616
/note="AluSc repeat: matches 1..297 of consensus"
complement(37235..37677)
/note="match: GSS: Em:A0058283"
38768..39332
/note="Cpg Island"
/evidence="not_experimental"
complement(39645..40061)
/note="match: STS: Em:G05971"
39646..39826
/note="match: STS: Em:G59992"
39759..39820
/note="2 copies 31 mer 100% conserved"
41211..41498
/note="AluSx repeat: matches 1..288 of consensus"
41830..41895
/note="MER91B repeat: matches 1..66 of consensus"
45630..45734
/note="MIR repeat: matches 148..256 of consensus"
45811..46027
/note="MIR repeat: matches 39..261 of consensus"
46028..46069
/note="14 copies 3 mer cat 100% conserved"
46245..46473
/note="match: STS: Em:HSPH32C9"
46546..46628
/note="MER33 repeat: matches 238..324 of consensus"
46629..46907
/note="AluSx repeat: matches 1..280 of consensus"
46908..47163
/note="MER33 repeat: matches 4..228 of consensus"
47214..47332
/note="MIR repeat: matches 12..137 of consensus"
47409..47662
/note="L1MB4 repeat: matches 5924..6181 of consensus"
47817..48194
/note="MIR repeat: matches 1..390 of consensus"
48463..48492
/note="15 copies 2 mer tt 90% conserved"
48511..48946
/note="match: STS: Em:G36957"
48722..48834
/note="MIR repeat: matches 35..138 of consensus"
49106..49234
/note="L2 repeat: matches 2605..2750 of consensus"
49628..50196
/note="L1PA16 repeat: matches 5577..6144 of consensus"
50428..50717
/note="AluSx repeat: matches 1..292 of consensus"
complement(50645..50888)
/note="match: GSS: Em:A0105432"
misc_feature

```

```

misc_feature 51632..52077

Query Match 30.2%; Score 475.4; DB 9; Length 58097;
Best Local Similarity 97.6%; Pred. No. 9..9e-100;
Matches 493; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 1051 ttacatcaaacaccctgtagagaagctttgcgcgtctctctaaatagctgcacgt 1110
|||||
DB 40137 ttacatcaaacaccctgtagagaagctttgcgcgtctctctaaatagctgcacgt 40078

QY 1111 ctttagaatgttaatgccccgattatcatcttgcctcgtatttaattactga 1170
|||||
DB 40077 ctttagaatgttaatgccccgattatcatcttgcctcgtatttaattactga 40018

QY 1171 ttccagaagaccacaatcatctgtatcagattatttaagtttacccgtagtttgata 1230
|||||
DB 40017 ttccagaagaccacaatcatctgtatcagattatttaagtttacccgtagtttgata 39958

QY 1231 aaagatttccatcttccctggctcgtcagagaacctaaagtgctacttgcaat 1290
|||||
DB 39957 aaagatttccatcttccctggctcgtcagagaacctaaagtgctacttgcaat 39898

QY 1291 ggcaagactagggtctatgtcttttaacccttaaaaaaatgttaagctgtatcc 1350
|||||
DB 39897 ggcaagactagggtctatgtcttttaaccctttt-AAAAAAAAATGTAAAGCTTAGTTACC 39839

QY 1351 tacttttcttgaatttcgacgtttgactagcaccatcaagaactttcgacgtttgac 1410
|||||
DB 39838 tacttttcttgaatttttgacgttttgacgttagcaccatcaagaactttcgacgtttgac 39779

QY 1411 tagcatctcaagaagtttaatcaagaatcatcagctgtatcattgactactca 1470
|||||
DB 39778 tagcatctcaagaagtttaatcaagaatcatcagctgtatcattgactactca 39719

QY 1471 acaaaaggaaagggtgtcagaagatcatatgaatttcgtccaaatttcaataatt 1530
|||||
DB 39718 acaaaaggaaagggtgtcagaagatcatatgaatttcgtccaaatttcaataatt 39659

QY 1531 tctcttctctcttaaaaaa 1555
|||||
DB 39658 tctcttctctctcttaaaaaa 39634

RESULT 12
LOCUS G05971 437 bp DNA linear STS 19-OCT-1995
DEFINITION human STS WI-7773, sequence tagged site.
ACCESSION G05971
VERSION G05971.1 GI:859216
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the Unigene collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 437)
REFERENCE
AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped ESTs
JOURNAL Unpublished
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: TTAATTTCGATTCACGACGAC
Primer B: TCTGACACCTTCCTTTG
STS size: 333
PCR Profile:

```

Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
Thermal Cycler:
Protocol:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul
Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCL: 10 mM
 pH: 9.3

Db	1 T 1	
RESULT 14		
AL137141		
LOCUS	180680 bp DNA linear PRI 24-MAY-2000	
DEFINITION	Human DNA sequence from clone RP11-139H14 on chromosome 13 contains the LCP1 (lymphocyte cytosolic protein 1 (L-plastin)) gene, part of the gene for CPB2 (carboxypeptidase B2 (plasma)), ESTs, STSs, GSSs and a CpG island, complete sequence.	
ACCESSION	AL137141	
VERSION	AL137141.10 GI:8217499	
KEYWORDS	HTG; carboxypeptidase B2; CPB2; CpG Island; cytosolic; L-plastin; LCP1; lymphocyte.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 180680)	
TITLE	Direct Submission	
JOURNAL	Submitted (24-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk	
COMMENT	<p>On Jun 3, 2000 this sequence version replaced gi:7523347. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.</p> <p>This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.</p> <p>The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SW., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr13</p> <p>RP11-139H14 is from the library RPCI-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR:PBAC3.6</p> <p>This sequence is the entire insert of clone RP11-139H14 The true left end of clone RP11-106H11 is at 103183 in this sequence.</p>	
FEATURES	location/Qualifiers	
source	1. 180680	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="13"	
	/clone="RP11-139H14"	
	/clone_11b="RPCI-11.1"	
misc_feature	453..898	
	/note="match: GSS: Em: A0601126"	
repeat_region	465..544	
	/note="40 copies 2 mer ta 75% conserved"	
repeat_region	548..754	
	/note="L1MC4 repeat: matches 6571..6787 of consensus"	
repeat_region	1297..1361	
	/note="MIR repeat: matches 82..146 of consensus"	
repeat_region	2531..2642	
	/note="L1M4 repeat: matches 4637..4744 of consensus"	
repeat_region	2674..3981	
	/note="L1M4 repeat: matches 2110..3419 of consensus"	
repeat_region	5917..6306	
	/note="MSTB repeat: matches 1..426 of consensus"	
repeat_region	6665..7452	
	/note="L1MC repeat: matches 2400..2858 of consensus"	
repeat_region	7453..7742	
	/note="AluYo repeat: matches 1..290 of consensus"	
repeat_region	7743..7940	
	/note="L1MC repeat: matches 2209..2400 of consensus"	
repeat_region	7971..8339	
	/note="L1MC repeat: matches 1855..2223 of consensus"	
misc_feature	8282..8613	
	/note="match: GSS: Em: A082919"	
repeat_region	8454..8548	
	/note="Alu repeat: matches 1..95 of consensus"	
repeat_region	8538..8761	
	/note="L1 repeat: matches 2193..2404 of consensus"	
repeat_region	8754..9754	
	/note="L1M1 repeat: matches 1156..2544 of consensus"	
repeat_region	9764..11803	
	/note="L1PA13 repeat: matches 4126..6144 of consensus"	
repeat_region	11804..12109	
	/note="AluY repeat: matches 1..307 of consensus"	
repeat_region	12110..13463	
	/note="L1PA13 repeat: matches 2741..4126 of consensus"	
repeat_region	13464..13816	
	/note="THE1B repeat: matches 1..364 of consensus"	
repeat_region	13817..14231	
	/note="L1PA13 repeat: matches 2332..2741 of consensus"	
repeat_region	14244..14735	
	/note="L1 repeat: matches 3036..3533 of consensus"	
repeat_region	14739..15072	
	/note="L1MC1 repeat: matches 5928..6260 of consensus"	
repeat_region	15087..16456	
	/note="L1MC repeat: matches 1271..2076 of consensus"	
repeat_region	16457..16559	
	/note="013 repeat: matches 1..103 of consensus"	
repeat_region	16560..17236	
	/note="L1MC repeat: matches 588..1271 of consensus"	
repeat_region	17243..17307	
	/note="L1MD3 repeat: matches 7682..7739 of consensus"	
repeat_region	17318..17480	
	/note="MIR repeat: matches 20..187 of consensus"	
repeat_region	19367..19524	
	/note="L1MCB repeat: matches 1032..1190 of consensus"	
repeat_region	19561..19592	
	/note="16 copies 2 mer ct 84% conserved"	
repeat_region	19945..20115	
	/note="MSTB repeat: matches 1..169 of consensus"	
repeat_region	20116..20414	
	/note="AluY repeat: matches 1..299 of consensus"	
repeat_region	20425..20725	
	/note="AluS repeat: matches 1..305 of consensus"	
repeat_region	20726..20964	
	/note="MSTB repeat: matches 177..426 of consensus"	
repeat_region	20966..21618	
	/note="MSTB-internal repeat: matches 2..667 of consensus"	
repeat_region	21629..21931	
	/note="MLT2D repeat: matches 233..553 of consensus"	
repeat_region	21932..22073	
	/note="MSTB-internal repeat: matches 1501..1650 of consensus"	
repeat_region	22075..22480	
	/note="MSTB repeat: matches 1..426 of consensus"	
repeat_region	22481..22589	
	/note="FLAM-A repeat: matches 1..113 of consensus"	
misc_feature	24229..24707	
	/note="match: STS: Em: T03078"	
repeat_region	25947..26255	
	/note="AluYo repeat: matches 1..311 of consensus"	
repeat_region	26647..26749	
	/note="L2 repeat: matches 1161..1267 of consensus"	
repeat_region	27045..28334	
	/note="L1M2 repeat: matches 413..1676 of consensus"	
repeat_region	28335..28656	
	/note="AluSg repeat: matches 1..310 of consensus"	

```

repeat_region 28657..28888
/note="LIM2 repeat: matches 1676..1910 of consensus"
repeat_region 28882..28967
/note="LIM1 repeat: matches 4801..4906 of consensus"
repeat_region 29015..29706
/note="LIM3 repeat: matches 5455..6146 of consensus"
repeat_region 29707..29918
/note="LIM1 repeat: matches 4892..5103 of consensus"
repeat_region 29924..30246
/note="LIM1 repeat: matches 5212..5403 of consensus"
repeat_region 31065..31296
/note="L2 repeat: matches 2477..2706 of consensus"
repeat_region 31301..37567
/note="L1P9 repeat: matches 144..6143 of consensus"
repeat_region 37667..37730
/note="MER20 repeat: matches 159..218 of consensus"
repeat_region 38334..38909
/note="MER82 repeat: matches 6..605 of consensus"
misc_feature 41615..42079
/note="match: GSS: Em: A0583743"
complement(42645..42669)
/note="match: STS: Em: G24043"
41129..43189
/note="L2 repeat: matches 2648..2708 of consensus"
repeat_region 43661..43874
/note="MIR repeat: matches 23..253 of consensus"
46480..46765
/note="AluJo repeat: matches 1..287 of consensus"
repeat_region 46766..46952
/note="L1 repeat: matches 5164..5370 of consensus"
47102..47417
/note="AluJo repeat: matches 1..311 of consensus"
repeat_region 47518..47759
/note="MIR repeat: matches 13..262 of consensus"
47887..47928
/note="21 copies 2 mer ta 83% conserved"
49439..49482
/note="MIR repeat: matches 107..150 of consensus"
49517..49757
/note="MIR repeat: matches 6..262 of consensus"
50473..50594
/note="L2 repeat: matches 545..679 of consensus"
50599..50898
/note="AluX repeat: matches 1..304 of consensus"
50906..51438
/note="LIMD3 repeat: matches 7223..7739 of consensus"
51480..52374
/note="L2 repeat: matches 693..1543 of consensus"
52375..52536
/note="L1B2 repeat: matches 5997..6155 of consensus"
52537..52577
/note="L2 repeat: matches 1543..1581 of consensus"
52578..52882
/note="AluX repeat: matches 1..305 of consensus"
52883..53971
/note="L2 repeat: matches 1581..2750 of consensus"
56191..56608
/note="MIR repeat: matches 8..426 of consensus"
57138..57237

```

Query Match 8.1% Score 128; DB 9; Length 180680;
 Best Local Similarity 87.5% Pred. No. 2.2e-19;
 Matches 140; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Oy 133 ttcaagttctacagattctactacacatgagatgtctctcgtgagccggttaag 192
 |||||
 Db 167852 TTTTATTTTTCATTTTGGCCCTTAACGTGATGTTCTCTGCGACGCGTAACAG 167911
 |||||

Oy 193 ctgacctattgtgagaaaaaagtcacattttttgttaaatgcatctgattgctgaca 252
 |||||
 Db 167912 CTGACCTATTGTGTAAGAAAAAAGTCACATTTTGTGTAATGATCATCTGATGACA 167971

Oy 253 atgtgaagccatttaaatgtgagcggaattcattgacg 292

```

|||||
Db 167972 ATGTGAAGCCCATTTAATGTAGCGGAATTCATGACG 168011

RESULT 15
LOCUS AR072911 927 bp DNA linear PAT 28-AUG-2000
DEFINITION Sequence 5 from patent US 5948668.
ACCESSION AR072911
VERSION AR072911.1 GI:9999674
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 927)
AUTHORS Hartman,J., Fulga,N., Mendelovitch,S. and Gorecki,M.
TITLE Production of enzymatically active recombinant carboxypeptidase B
JOURNAL Patent: US 5948668-A 5 07-SEP-1999;
FEATURES
source location/Qualifiers
1..927
/organism="unknown"
BASE COUNT 253 a 226 c 213 g 235 t
ORIGIN

```

Query Match 7.9% Score 124; DB 6; Length 927;
 Best Local Similarity 57.9% Pred. No. 2.3e-18;
 Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

Oy 613 gaatgtgagaagaagacggtcttctatctgacacatcatctcaggaagacactga 672
 |||||
 Db 362 GAATGTGAGAAAGAACCCGCTCTACATGCTGGAGTTCTCTGGGTGATGACCCCA 421

Oy 673 atagcaattgtctccaacactggtgtgaggaaggtgcatcagttccatgtctcg 732
 |||||
 Db 422 ACAGCAATTTTAATGCG--TGCTGTGTGAGAGTGGAGGCTTCTCGAGTCCCTGCTG 478

Oy 733 aaactactgtgacttaccctcagagtcagagacagaagtggaagcgagtgtagttct 792
 |||||
 Db 479 AAACCTTCTGTGGACGACGCCGACAGTCTGTAAGAAAGGACAAAGGCCCTGCGAGATTTC 538

Oy 793 tgaagaagaatlcacacagattaaagacalacalcagcatgcatlactccagcata 852
 |||||
 Db 539 TCGCACAACACCTCTCCACATCAAGGCTACCTGACCATCTCATCTACACAGATGA 598

Oy 853 taagtctcatatctctatatacagaagtaagaagcaagccatgaggaactgtctag 912
 |||||
 Db 599 TGCTTACCTTACTCTTACTATGACTACAACTGCTGGAATGAGAAATGAAATGCCC 658

Oy 913 tagcagtgagagcagtcgctgctatgtacaaaactagtaaaataccaggtatcacatg 972
 |||||
 Db 659 TGTGGAAGGTGCGGCAAGAGAGCTTCC--ACTGTGCAATGGGACCAAGTACATATG 715

Oy 973 gccatgctcagaaaccttaacctagctcctctgaggttgaggagcatgatatgatt 1032
 |||||
 Db 716 GCCAGAGAGTACAAACATCTATCTCTGCTGGGGATCTGACGACTGGTCTTATATC 775

Oy 1033 tgggcataaatatgtgttaccatcaaacacacacgtgagagaagacttgccggtctc 1092
 |||||
 Db 776 AGGGAATCAAAATTCCTTACCTTTGA-ACCTCCGGATACAGGCTTTTGGCTTCTC 834

Search completed: September 18, 2002, 16:05:39
 Job time: 2842 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2002, 15:24:02 ; Search time 210.94 Seconds

(without alignments)
12803.197 Million cell updates/sec

Title: US-09-980-881-1

Perfect score: 1573
Sequence: 1 agaaaaatcgttgggatg.....aaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802:*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1573	100.0	1573	21	AAC81962 Human brain carbox
2	1372	87.2	1749	17	AAT11671 Human plasma carbo
3	1372	87.2	1749	18	AAT62846 Human plasma carbo
4	1370.4	87.1	1749	14	AAQ41001 Human plasma carbo
5	918	58.4	1272	20	AAV74302 Human plasma carbo
6	853.8	54.3	1400	21	AAFI8005 Lung cancer associ
7	324	20.6	333	20	AAH85957 Human single nucle
8	324	20.6	333	20	AAH87390 Human single nucle
9	241.6	15.4	251	19	AAH11374 Human biallelic po

10	124	7.9	927	17	AAT35760 Rat mature carboxy
11	123.4	7.8	921	16	AAQ90601 Porcine carboxypep
12	123.4	7.8	1215	16	AAQ90600 Porcine Tyr-His-Me
13	120	7.6	121	20	AAH85710 Human single nucle
14	120	7.6	121	20	AAH85711 Human single nucle
15	109	6.9	1263	19	AAV41795 Human pancreatic c
16	107.4	6.8	999	17	AAT42500 Mature HCPB coding
17	107.4	6.8	1053	17	AAT42497 Mature HCPB-(His)6
18	107.4	6.8	1263	17	AAT42494 Human pancreatic c
19	107.4	6.8	1284	17	AAT42506 ProHCPB gene with
20	104.2	6.6	1053	20	AAH24804 Human carboxypept
21	104.2	6.6	1059	17	AAT42511 Modified HCPB (D25
22	104.2	6.6	1059	17	AAT42512 Modified HCPB (D25
23	104.2	6.6	1059	18	AAT62787 Carboxypeptidase B
24	104.2	6.6	1059	18	AAT62788 Carboxypeptidase B
25	102.6	6.5	1059	18	AAT62805 Carboxypeptidase B
26	102.6	6.5	1059	18	AAT62806 Carboxypeptidase B
27	102.6	6.5	1059	18	AAT62791 Carboxypeptidase B
28	102.6	6.5	1059	18	AAT62801 Carboxypeptidase B
29	101	6.4	1059	18	AAT62789 Carboxypeptidase B
30	101	6.4	1059	18	AAT62790 Carboxypeptidase B
31	101	6.4	1870	18	AAV17322 DNA encoding human
32	101	6.4	1870	19	AAV17322 Human carboxypepti
33	101	6.4	2154	18	AAV17331 Carboxypeptidase B
34	99.4	6.3	1059	18	AAT62808 Carboxypeptidase B
35	99.4	6.3	1059	18	AAT62809 Carboxypeptidase B
36	99.2	6.3	1059	18	AAT62800 Carboxypeptidase B
37	97.8	6.2	1059	18	AAT62807 Carboxypeptidase B
38	97.8	6.2	1059	18	AAT62803 Carboxypeptidase B
39	96.2	6.1	1059	18	AAT62802 Carboxypeptidase B
40	92.8	5.9	1311	22	AAD03837 Human carboxypepti
41	92.8	5.9	2128	22	AAH34774 Human colon cancer
42	91.6	5.8	936	22	AAH34774 Human breast cancer
43	88.6	5.6	527	22	AAI22383 Human carboxypepti
44	88.6	5.6	1050	22	AAH03838 Human carboxypepti
45	87.6	5.6	1200	22	AAH05477 Novel human protea

ALIGNMENTS

RESULT 1

AAH81962 standard; cDNA; 1573 BP.

ID AAC81962;

AC AAC81962;

AC 01-MAR-2001 (first entry)

DT Human brain carboxypeptidase B cDNA.

DE Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase; treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine; cerbroprotective; antilzheimer's; neuroprotective;

KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;

KW Down's syndrome; head trauma; ss.

KW Homo sapiens.

OS Key

FT Location/Qualifiers

FT CDS 18..1100

FT /tag= a

FT /product= "carboxypeptidase B"

PN WO200066717-A1.

XX 09-NOV-2000.

XX 01-MAY-2000; 2000WO-JP02878.

XX 30-APR-1999; 99JP-0125169.

XX (MATS/) MATSUMOTO A.

XX Matsumoto A;
XX WPI: 2000-687534/67.
DR P-PSDB: AAB11457.
XX Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Claim 2b; Page 64-68; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimer's,
CC neurotropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;

Query Match 100.0%; Score 1573; DB 21; Length 1573;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaaaattgctgttggatggaagcttgcagccttgacagcttgtaaccattgttctct 60
Db 1 agaaaattgctgttggatggaagcttgcagccttgacagccttgtaaccattgttctct 60
QY 61 tctgtgagcagcagcttcttcgcttccagagtgagcgaagtcctcagctcttcctgaa 120
Db 61 tctgtgagcagcagcttcttcgcttccagagtgagcgaagtcctcagctcttcctgaa 120
QY 121 cctctagcagaagtcagaattctacagaattcttactacacatagagattgtctcggc 180
Db 121 cctctagcagaagtcagaattcttactacacatagagattgtctcggc 180
QY 181 agccggttaacagcttgacattatgtgaagaaaaacaagtcattttttgttaaatgcat 240
Db 181 agccggttaacagcttgacattatgtgaagaaaaacaagtcattttttgttaaatgcat 240
QY 241 ctgagtgcgacaagtgtgaagccatttaattgtgagcgaattccatgagtgcttcgc 300
Db 241 ctgagtgcgacaagtgtgaagccatttaattgtgagcgaattccatgagtgcttcgc 300
QY 301 tggcagagcgtggaagatctatcaacagagaatttccacagacagctcagcccgag 360
Db 301 tggcagagcgtggaagatctatcaacagagaatttccacagacagctcagcccgag 360
QY 361 ccttcgcagctgactatgaacagatcaactcaaatgaattatcttctgtgataaat 420
Db 361 ccttcgcagctgactatgaacagatcaactcaaatgaattatcttctgtgataaat 420
QY 421 ttataacttgagagcagctcgtatgttcaaaaaatccagattggtcttcatttga 480
Db 421 ttataacttgagagcagctcgtatgttcaaaaaatccagattggtcttcatttga 480
QY 481 agtaccacactctatgttttaaggttctctgaaaagaaacaacgcaaaaatgacctat 540
Db 481 agtaccacactctatgttttaaggttctctgaaaagaaacaacgcaaaaatgacctat 540
QY 541 ggaattgactgtgaaatccatgacagagaatgagatctcctgcttctgtgtgttca 600
Db 541 ggaattgactgtgaaatccatgacagagaatgagatctcctgcttctgtgtgttca 600
QY 601 taggcataatcgaaatgtggaagaaacggtcttctctatgtaaaatcatgtgctcg 660
Db 601 taggcataatcgaaatgtggaagaaacggtcttctctatgtaaaatcatgtgctcg 660

QY 661 gaacagacctgaatagcaacttgtctccaacacatggtggaaggaagtgcaccagtt 720
Db 661 gaacagacctgaatagcaacttgtctccaacacatggtggaaggaagtgcaccagtt 720
QY 721 cctcatgtcggaaacctactcgtgacttatacctgaatcagaacccaagatggaagcag 780
Db 721 cctcatgtcggaaacctactcgtgacttatacctgaatcagaacccaagatggaagcag 780
QY 781 tggctagtcttcttgagaagaataccaacagattaaagcatcaatcagcatcatcat 840
Db 781 tggctagtcttcttgagaagaataccaacagattaaagcatcaatcagcatcatcat 840
QY 841 actccagacatagtgttccatattctctatacagaagaagaaagaacacatagag 900
Db 841 actccagacatagtgttccatattctctatacagaagaagaaagaacacatagag 900
QY 901 aactgctcctagtagcagatgaagcagcttgcgtatgtgaacaaactagtaaaatacca 960
Db 901 aactgctcctagtagcagatgaagcagcttgcgtatgtgaacaaactagtaaaatacca 960
QY 961 ggtatacaatgagcagatggtcagaaacattatacctagctcctgaggtgaggagatt 1020
Db 961 ggtatacaatgagcagatggtcagaaacattatacctagctcctgaggtgaggagatt 1020
QY 1021 ggaatcatgatttggcacaataatcgtttacatacaaaccccggtgagaagattt 1080
Db 1021 ggaatcatgatttggcacaataatcgtttacatacaaaccccggtgagaagattt 1080
QY 1081 tgcgcgtgctcttaaatagcttgcagatgcatataggaagatttaagccocgatttta 1140
Db 1081 tgcgcgtgctcttaaatagcttgcagatgcatataggaagatttaagccocgatttta 1140
QY 1141 tcatttcgcttcgattttaaattactgattccagaagaaccaatcatgtatcagat 1200
Db 1141 tcatttcgcttcgattttaaattactgattccagaagaaccaatcatgtatcagat 1200
QY 1201 tatttttaagttatccgtatgttgaataaagaatttccatctccttggttcgcag 1260
Db 1201 tatttttaagttatccgtatgttgaataaagaatttccatctccttggttcgcag 1260
QY 1261 agaacctaataagtgctactcttgccatgaagcagactaggttcacatgcttttaacct 1320
Db 1261 agaacctaataagtgctactcttgccatgaagcagactaggttcacatgcttttaacct 1320
QY 1321 ttaaaaaaaattgttaaaagtctgaactactcttcttcttgatctcgaggttgaact 1380
Db 1321 ttaaaaaaaattgttaaaagtctgaactactcttcttcttgatctcgaggttgaact 1380
QY 1381 agccatctcaagcaacttgcagcttgcactgagcagcatctcaagcaagtttaacaagaat 1440
Db 1381 agccatctcaagcaacttgcagcttgcactgagcagcatctcaagcaagtttaacaagaat 1440
QY 1441 catctcagctgcatcatgtgatacctactcaacaaaagaaggtgycagagaagtatactta 1500
Db 1441 catctcagctgcatcatgtgatacctactcaacaaaagaaggtgycagagaagtatactta 1500
QY 1501 aagattctcgtccaaatttcaataaatttcttctcctttaaataaaaaaa 1560
Db 1501 aagattctcgtccaaatttcaataaatttcttctcctttaaataaaaaaa 1560
QY 1561 aaaaaaaaaa 1573
Db 1561 aaaaaaaaaa 1573

RESULT 2
AAT11671
ID AAT11671 standard; DNA; 1749 BP.
XX
AC AAT11671;
XX
DT 12-APR-1996 (first entry)
XX

DE Human plasma carboxypeptidase B coding sequence.
XX
KW Plasma carboxypeptidase B; hpcpb; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 41..1312
FT CDS /tag= a
FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT /tag= b
FT mat_peptide 107..1309
FT /tag= c
XX
PN US5474901-A.
XX
XX 12-DEC-1995.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 14-OCT-1992; 92US-0959944.
XX 15-DEC-1993; 93US-0167727.
XX 19-JUL-1994; 94US-0277540.
XX
XX (GENTH) GENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI; 1996-039508/04.
XX P-PSDB; AAR90293.
XX
XX Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hpcpb for use in treating clotting disorders e.g.
PT haemophilia A
XX
XX Disclosure; Figure 4; 40pp; English.
XX
XX An antibody which specifically binds human plasma carboxypeptidase B
CC (hpcpb) and does not cross react with other carboxypeptidases is
CC useful for the detection of hpcpb in vitro. The antibody is also
CC used for purifying hpcpb from a sample. Purification comprises
CC passing a sample thought to contain hpcpb over either a column to
CC which antibody has been bound, or a plasminogen affinity column,
CC eluting the column and then recovering the fraction containing the
CC hpcpb.
XX
XX Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;
XX

Db 264 ctgattgcacaaatgtaagccatttaattgtagcggaattccatgcatgctctgc 323
QY 301 tggcaaacgtggaagattctattccaacagaagattccacagacagtcagcccgag 360
Db 324 tggcagacgtggaagattctattccaacagagattccacagacagtcagcccgag 383
QY 361 cctcgcatcgtactatgaacagatcactactaaatgaatctattcttgatgaat 420
Db 384 cctccgcacgtactatgaacagatcactactaaatgaatctattcttgatgaat 443
QY 421 ttataactgagagcgatcctgattatgcttaacaaaatccatttgatcttattgaa 480
Db 444 ttataactgagagcgatcctgattatgcttaacaaaatccatttgatcttattgaa 503
QY 481 agtaccactctatgttttaaaagttctggaagaagaacacagcaaaaatgcatat 540
Db 504 agtaccactctatgttttaaaagttctggaagaagaacacagcaaaaatgcatat 563
QY 541 ggattgaactggaatccatgccaagaatggaatcctcctgcttctgcttggttca 600
Db 564 ggattgaactggaatccatgccaagaatggaatcctcctgcttctgcttggttca 623
QY 601 taggcacat----- 608
Db 624 taggcacataactcaattctatggaataagggcaataataccaaatcctctgagcttg 683
QY 609 ----- 609
Db 684 tggattctcatgttatgcccgggtggttaatgtagcggtaatactcactcatggaaaaaga 743
QY 610 atcgaaatgtagaagaacacggttcttctatgtagaacaatcatgcatcggaacagacc 669
Db 744 atcgaaatgtagaagaacacggttcttctatgtagaacaatcatgcatcggaacagacc 803
QY 670 tgaatgcaacttgcctcacaacactggtgtagaagaatgtagatccattctctatgct 729
Db 804 tgaatgcaacttgcctcacaacactggtgtagaagaatgtagatccattctctatgct 863
QY 730 cggaaacctactgtgacttactcctgagtcgaacgaacgaagtagagcgatgagct 789
Db 864 cggaaacctactgtgacttactcctgagtcgaacgaacgaagtagagcgatgagct 923
QY 790 tcttgaagaagaatatacaaccagattaaagcatatcacatgcatgcatctactccagc 849
Db 924 tcttgaagaagaatatacaaccagattaaagcatatcacatgcatgcatctactccagc 983
QY 850 atatagtgcttcataattcctatacagaagtaagaagcaagcattgaggactgtctc 909
Db 984 atatagtgcttcataattcctatacagaagtaagaagcaagcattgaggactgtctc 1043
QY 910 tagtaacagtagaagatctggtcatttggaacaaactagtaaaaatacagaattacac 969
Db 1044 tagtaacagtagaagatctggtcatttggaacaaactagtaaaaatacagaattacac 1103
QY 970 atggccatggtctcagaacattatcctagctctcgtgaggtgggaagatlgatctatg 1029
Db 1104 atggccatggtctcagaacattatcctagctctcgtgaggtgggaagatlgatctatg 1163
QY 1030 atttggcatcaaatatgt----- 1050
Db 1164 atttggcatcaaatatgt----- 1223
QY 1051 -----ttacacaaacacacgtgtagagaacgtttgcccgtgctcctcaaaa 1097
Db 1224 tgcgtgcggaagcgttacacaaacacacgtgtagagaacgtttgcccgtgctcctcaaaa 1283
QY 1098 tagcttgcatgtcatatgaagaatggttaatgcccgtgatttatacattcgtctccgcat 1157
Db 1284 tagcttgcatgtcatatgaagaatggttaatgcccgtgatttatacattcgtctccgcat 1343
QY 1158 tttaattctgattctcagaagaacaaatcattgataagattattttaagtttattc 1217

Db	1344	tttaattactgattccgaagaaaccaaatcatgtatcagttatttttaagttttac	1409
Qy	1218	cgtaagttttgataaagaattttccattctcttggtgcgtgtcagagaacctaaatgagtc	1277
Db	1404	cgtaagttttgataaagaattttccattctcttggtgcgtgtcagagaacctaaatgagtc	1463
Qy	1278	aattcgcatataagcagagactaggtgtctgtcttttcccttttaaaaaaatgttaa	1337
Db	1464	aattcgcatataagcagagactaggtgtctgtcttttcccttttaaaaaaatgttaa	1523
Qy	1338	aagctcgtttacactacttttctcttgatttttcgaagctttgactagcatctcaagaact	1397
Db	1524	aagctcgtttacactacttttctcttgatttttcgaagctttgactagcatctcaagaact	1583
Qy	1398	ttcgaagctttgactagcatctcaagaactttaatcaagaatcatctcaagctgcatcat	1457
Db	1584	ttcgaagctttgactagcatctcaagaactttaatcaagaatcatctcaagctgcatcat	1643
Qy	1458	tggatcctactcaacaaagaaggaaggtggtctcagaagtatcattaagaattctgtctcaaa	1517
Db	1644	tggatcctactcaacaaagaaggaaggtggtctcagaagtatcattaagaattctgtctcaaa	1703
Qy	1518	ttttcaataaattctctctctcccttttaaaaaaaaaaaaaaaaaa	1563
Db	1704	ttttcaataaattctctctctcccttttaaaaaaaaaaaaaaaaaa	1749

RESULT	3
AAAT62846	
ID	AAAT62846 standard; DNA; 1749 BP.
XX	
AC	AAAT62846;
XX	
DT	08-MAY-1997 (first entry)
XX	
DE	Human plasma carboxypeptidase B coding sequence.
XX	
KW	Human; plasma carboxypeptidase B; PCPB; hemostatic regulation;
XX	
XX	plasma; plasminogen; ss.
OS	Homo sapiens.
XX	
FH	key
FT	location/qualifiers
FT	41..1312
FT	/*tag= a
FT	/product= Human PCPB
FT	41..106
FT	/*tag= b
FT	107..1309
FT	/*tag= c
FT	134..177
FT	/*tag= d
FT	/bound_moety= 46_bp_probe
XX	
PN	US5593674-A.
XX	
PD	14-JAN-1997.
XX	
PF	01-FEB-1991; 91US-0649591.
XX	
PR	01-FEB-1991; 91US-0649591.
PR	14-OCT-1992; 92US-0959944.
PR	15-DEC-1993; 93US-0167727.
PR	19-JUL-1994; 94US-0277540.
PR	27-APR-1995; 95US-0430787.
XX	
PA	(GETH) GENENTECH INC.
PI	Drayna DT, Eaton DL;
XX	
PI	WPI; 1997-099413/09.
DR	P-PSDB; AAAT4733.
XX	

PT Using human plasma carboxy:peptidase B in blood coagulation - 1s
 PT functionally related to carboxy:peptidase A and pancreas
 PT carboxy:peptidase B
 XX
 PS
 XX Example 2; Column 37-42; 39pp; English.
 CC
 CC This sequence encodes human plasma carboxypeptidase B (PCPB) which
 CC has a molecular weight under non-reducing SDS-PAGE of approx. 60kD
 CC PCPB may be used therapeutically in haemostatic regulation.
 CC purified from human plasma or by transmembrane cell culture by
 CC extraction using plasminogen bound to a solid phase.
 XX
 S0 Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Query Match	87.2%;	Score 1372;	DB 18;	Length 1749;
Best Local Similarity	90.3%;	Pred. No. 0;		
Matches 1558; Conservative	0;	Mismatches 5;	Indels 163;	Gaps 2;

QY	61	tccttgcagcaacgagctcttcgcgtctccagagctggccagattctcagcttcgctcccttcaggaa	120
Db	84	tccttgagcagcagctctctcgtctcgtttccagagctggccagattctcagcttcgctcccttcaggaa	143
QY	121	ccctcaggaagaattccaaagttctctacagaaactcttactacaacatcagagatgtgtctctcgc	180
Db	144	ccctctcaggaagaattccaaagttctctacagaaactcttactacaacatcagagatgtgtctctcgc	203
QY	181	agcccgtaacagcgcgactcttctgtgagaagaaacaagccatctttttgtgataatgcat	240
Db	204	agcccgtaacagcgcgactcttctgtgagaagaaacaagccatctttttgtgataatgcat	263
QY	241	ctgactgtcgaacacatgtgtgaaagcccatcttaaatgtgtgagcgggaattccatctgcagctgtctgc	300
Db	264	ctgactgtcgaacacatgtgtgaaagcccatcttaaatgtgtgagcgggaattccatctgcagctgtctgc	323
QY	301	tggcagaagcgtgtgaaagatctctatttacaacagcagatcttccaacgcagacagtcagcccccag	360
Db	324	tggcagaagcgtgtgaaagatctctatttacaacagcagatcttccaacgcagacagtcagcccccag	383
QY	361	ccctccgacatcgtactatgcatacagcatcatccacacaaatgaaatccatctctctgtgatagaat	420
Db	384	ccctccgacatcgtactatgcatacagcatcatccacacaaatgaaatccatctctctgtgatagaat	443
QY	421	ttataactgagagagcgcacctcgtgatatgtcttaacaanaatccacatgttgatctctcatcttggaga	480
Db	444	ttataactgagagagcgcacctcgtgatatgtcttaacaanaatccacatgttgatctctcatcttggaga	503
QY	481	agttaccacactctatgtttttaagaagttctctcgtgaaaaagaacaacagccaaaaatgtgcatat	540
Db	504	agttaccacactctatgtttttaagaagttctctcgtgaaaaagaacaacagccaaaaatgtgcatat	563
QY	541	ggattgactgtgataatccatgcgcgcgagaaatgagatctctcctgtcttctcgttgtgttca	600
Db	564	ggattgactgtgataatccatgcgcgcgagaaatgagatctctcctgtcttctcgttgtgttca	623
QY	601	taggcgat-----	608
Db	624	taggcgatataaactcaattctatctatggtgataatagggcataatacacaatctcctgagcttg	683
QY	609	-----a	609
Db	684	tgcgaattctatgttatacgcgcggtgtgttatacgtgtgacggttatgactactccatctgaaaaaga	743
QY	610	atcgaaatgtgtgaaaaagaacacggtctctcttctatcgcgaacaacatcatctcatctcggagaacgc	669
Db	744	atcgaaatgtgtgaaaaagaacacggtctctcttctatcgcgaacaacatcatctcatctcggagaacgc	803
QY	670	tgaatagcaactctgtctccaaaacactgtgtgtgaggaaggtgcaccaggtctccatgct	729


```

Db 804 tgaataggaacttctccaaacacgtgtgtaggaaggtgacaccagttctcatgct 863
Oy 730 cggaaacctactgtggaacttctctgagtcagaaccagaagtgaagcagtgtagt 789
Db 864 cggaaacctactgtggaacttctctgagtcagaaccagaagtgaagcagtgtagt 923
Oy 790 tcttgaagaagaatctcaaccagatlaaagcatalcatagcgtatctatctccagc 849
Db 924 tcttgaagaagaatctcaaccagatlaaagcatalcatagcgtatctatctccagc 983
Oy 850 ataatgttctccatctctctacacgaagtaaaagcaaacatctgaggaactgtcc 909
Db 964 ataatgttctccatctctctacacgaagtaaaagcaaacatctgaggaactgtcc 1043
Oy 910 tagtagcagtaggaagcagttcgtctatgtgacaacactaglaaataatcacaggtatcac 969
Db 1044 tagtagcagtaggaagcagttcgtctatgtgagaacactaglaaataatcacaggtatcac 1103
Oy 970 atggccatggtcagaacactatactcctagctcgtgaggtgaggagcagattgattctatg 1029
Db 1104 atggccatggtcagaacactatactcctagctcgtgaggtgaggagcagattgattctatg 1163
Oy 1030 atttggcacaataatctcgt----- 1050
Db 1164 atttggcacaataatctcgttacaattgaaactcgagatacgggacacatcgattct 1223
Oy 1051 -----ttacatcaaacaccacgtgtagagaagcttttgcgcgtgctctaa 1097
Db 1224 tgcgtcgagagcgttaccatcaaacaccacgtgtagagaagcttttgcgcgtgctctaa 1283
Oy 1098 tagcttgacatgcatlaagaaatgtttaatgcccctgatttatacatcttctcctgcat 1157
Db 1284 tagcttgacatgcatlaagaaatgtttaatgcccctgatttatacatcttctcctgcat 1343
Oy 1158 ttaattactactatccagcagaacaaatctgtatcagaattattttaagttttac 1217
Db 1344 ttaattactactatccagcagaacaaatctgtatcagaattattttaagttttac 1403
Oy 1218 cgtagtttgaataaagaatttctctatctctgtgtctgctcagagaaactaataagtgct 1277
Db 1404 cgtagtttgaataaagaatttctctatctctgtgtctgctcagagaaactaataagtgct 1463
Oy 1278 acttggcatttaagcagactaggggtcgtcttcttaccctttaaaaaaattgttaa 1337
Db 1464 acttggcatttaagcagactaggggtcgtcttcttaccctttaaaaaaattgttaa 1523
Oy 1338 aagtcagttactcttcttcttgaatttgaagtttgaagttgactagacatcaagaact 1397
Db 1524 aagtcagttactcttcttcttgaatttgaagtttgaagttgactagacatcaagaact 1583
Oy 1398 ttcgacggttgaactagcactcacaagcaagtttaatacaagaatcatctcagctgcat 1457
Db 1584 ttcgacggttgaactagcactcacaagcaagtttaatacaagaatcatctcagctgcat 1643
Oy 1458 tggatccctactcaacaaagaaggtgtgtcagaagtaacatlaaagaatttctgtctcaaa 1517
Db 1644 tggatccctactcaacaaagaaggtgtgtcagaagtaacatlaaagaatttctgtctcaaa 1703
Oy 1518 ttttaataaattcttctctcctttaaaaaa 1563
Db 1704 ttttaataaattcttctctcctttaaaaaa 1749

```

RESULT 4
AA041001
ID AA041001 standard; cDNA; 1749 BP.

AC AA041001;
XX 24-AUG-1993 (first entry)
XX Human plasma carboxypeptidase B gene.
DE
XX

```

KW PCPB: tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH sig_peptide 41..106
FT /tag= a
FT mat_peptide 107..1312
FT /tag= b
FT /product= PCPB
FT misc_feature 134..177
FT /tag= C
FT /function= probe
FT /note= "used to obtain full-length clones"
PD US5206161-A.
XX 27-APR-1993.
PF 01-FEB-1991; 91US-0649591.
XX 01-FEB-1991; 91US-0649591.
XX 01-FEB-1991; 91US-0649591.
XX (GENTECH) GENENTECH INC.
XX Drayna DT, Eaton DL;
XX WPI; 1993-151724/18.
DR P-PDB; AAR36273.
XX
XX New human plasma carboxypeptidase B - used as haemostatic
PT regulator for clotting blood, partic. for treating blood clotting
PT disorders, e.g. haemophilia
XX
XX PS Disclosure; Fig 4; 40pp; English.
XX
XX Human plasma carboxypeptidase B was isolated from human plasma and
CC partially sequenced. Oligonucleotide primers were designed based on
CC the partial amino acid sequences. The primers were used in a PCR
CC amplification to identify cDNA encoding PCPB from a human liver cDNA
CC library. The PCR product was capable of encoding the first 37 amino
CC acids of PCPB; a 46mer probe was used to obtain the full-length
CC sequence which, although disclosed in the specification, is not
CC claimed. PCPB inhibits the enzymatic conversion by tPA of
CC plasminogen to plasmin in the presence of fibrinogen.
XX
SQ Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

```

Query Match 87.1%; Score 1370.4; DB 14; Length 1749;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 6; Indels 163; Gaps 2;

```

Oy 1 agaaattgctgttgatgaagaacttgcagccttgacgctctgtaaccattgtctct 60
Db 24 agaaattgctgttgatgaagaacttgcagccttgacgctctgtaaccattgtctct 83
Oy 61 tctgtagcagcagtgcttctgcgttcacagatggtgccaagttcagtgctctccctcaga 120
Db 84 tctgtagcagcagtgcttctgcgttcacagatggtgccaagttcagtgctctccctcaga 143
Oy 121 cctctaggaagttcaagttctacagaatcttactacaacatagatgttctctgagc 180
Db 144 cctctaggaagttcaagttctacagaatcttactacaacatagatgttctctgagc 203
Oy 181 agccggtacagtgacgtcttatgttgaagaagaagaagaagtcatttttgaatagcat 240
Db 204 agccggtacagtgacgtcttatgttgaagaagaagaagaagtcatttttgaatagcat 263
Oy 241 ctgattcgacaatgtgaaagccatttaaatgtgacggaatccatgacgtgtctgc 300
Db 264 ctgattcgacaatgtgaaagccatttaaatgtgacggaatccatgacgtgtctgc 323

```

```

QY 301 tggcagacgtggaagatcttatacaacagcagattccaacgacacagtcagcccccag 360
    |||
Db 324 tggcagacgtggaagatcttatacaacagcagattccaacgacacagtcagcccccag 383
QY 361 ccttcgacgtgtaatagaagatatacctacttaaatgaatacttcttggatagat 420
    |||
Db 384 ccttcgacgtgtaatagaagatatacctacttaaatgaatacttcttggatagat 443
QY 421 ttaatactgaagcagcctctgatactgcttaacaaatccaatctgacccatttgga 480
    |||
Db 444 ttaatactgaagcagcctctgatactgcttaacaaatccaatctgacccatttgga 503
QY 481 agtaccacactatgctttaaagttcttggaagaagaacaaacagcaaaatgccaat 540
    |||
Db 504 agtaccacactatgctttaaagttcttggaagaagaacaaacagcaaaatgccaat 563
QY 541 ggaatgacgtggaatcagtcagagaatgagctcctgccttctgcttggttca 600
    |||
Db 564 ggaatgacgtggaatcagtcagagaatgagctcctgccttctgcttggttca 623
QY 601 taaggcat----- 608
    |||
Db 624 taaggcatataacccaattctatggaatagggcaatatccaatctccttgagcttg 683
QY 609 -----a 609
    |||
Db 684 tgaattctatgctatgcccgttggttaatgtagcaggtatgactactatgaaaga 743
QY 610 atcgaatctggaagaagaacagctcttcttctatgagaacatcatctgagaaagacc 669
    |||
Db 744 atcgaatctggaagaagaacagctcttcttctatgagaacatcatctgagaaagacc 803
QY 670 tgaatacgaactctgctccaacacactggtggaagaagtgcacccagttcctatgct 729
    |||
Db 804 tgaatacgaactctgctccaacacactggtggaagaagtgcacccagttcctatgct 863
QY 730 cggaaacctactgtggaacttatccctgagtcagaacacgaagtgaaggcagtgtagct 789
    |||
Db 864 cggaaacctactgtggaacttatccctgagtcagaacacgaagtgaaggcagtgtagct 923
QY 790 tcttgaagaagaatatacacaacagattaaagatatacagatcatctacccagc 849
    |||
Db 924 tcttgaagaagaatatacacaacagattaaagatatacagatcatctacccagc 983
QY 850 atatagtcttccatctcctatacagcgaagtaaaagcaaaagacatlgaggaaactgtctc 909
    |||
Db 984 atatagtcttccatctcctatacagcgaagtaaaagcaaaagacatlgaggaaactgtctc 1043
QY 910 taagtagcagtggaagcagcttctgctatctgacaaactagtaaaataccaggtatatac 969
    |||
Db 1044 taagtagcagtggaagcagcttctgctatctgacaaactagtaaaataccaggtatatac 1103
QY 970 atggccatggtcagaaacctatacactagctcctgagagtgaggagacagatggatctatg 1029
    |||
Db 1104 atggccatggtcagaaacctatacactagctcctgagagtgaggagacagatggatctatg 1163
QY 1030 atttggcatcaaatatcg----- 1050
    |||
Db 1164 atttggcatcaaatatcg----- 1223
QY 1051 -----ttataataaacccactgtagaagaagctttggccgctgtccttaaa 1097
    |||
Db 1224 tggctgcgagcagcttatacaaacacccctgtagaagaagctttggccgctgtccttaaa 1283
QY 1098 taagctgacatgcatatagaagtgttaatgcccctgatttatacatctgctccgctat 1157
    |||
Db 1284 taagctgacatgcatatagaagtgttaatgcccctgatttatacatctgctccgctat 1343
QY 1158 tttaattactgattccaagaagaacaaatcatgtatcagatatttttaagtattatc 1217
    |||
Db 1344 tttaattactgattccaagaagaacaaatcatgtatcagatatttttaagtattatc 1403
QY 1218 cgtagtttataaagaatttctctatctccttggtctgtcagaagaacataaagtgtct 1277

```

```

Db 1404 cgtagtttataaagaatttctctatctccttggtctgtcagaagaacataaagtgtct 1463
    |||
QY 1278 atttggcatcaaatgaagcagactaggttcatgtcttttaacctttaaaaaaaattgttaa 1337
    |||
Db 1464 atttggcatcaaatgaagcagactaggttcatgtcttttaacctttaaaaaaaattgttaa 1523
QY 1338 aagctagtaacctactttcttcttgatcttcgaagtttgactagcactccaagaacct 1397
    |||
Db 1524 aagctagtaacctactttcttcttgatcttcgaagtttgactagcactccaagaacct 1583
QY 1398 ttgcagctttgactagcactccaagaagtttaatacagaatcatctcaagctgatacat 1457
    |||
Db 1584 ttgcagctttgactagcactccaagaagtttaatacagaatcatctcaagctgatacat 1643
QY 1458 tggatcctactccaagaagaaggttggtcaagaatgataaagatttctgctccaaa 1517
    |||
Db 1644 tggatcctactccaagaagaaggttggtcaagaatgataaagatttctgctccaaa 1703
QY 1518 ttccaataaatttcttctcctcttaaaaaaaataaaaaa 1563
    |||
Db 1704 ttccaataaatttcttctcctcttaaaaaaaataaaaaa 1749

```

```

RESULT 5
AAV74302
ID AAV74302 standard; cDNA; 1272 BP.
XX
AC AAV74302;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KM polymorphism detection; thrombotic disease; ds.
XX
OS Homo sapiens.
XX
PN M09855645-A1.
XX
PD 10-DEC-1998.
XX
PE 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD ) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
DR WPI: 1999-045800/04.
XX
P-PSDB: AAM92270.
XX
PT Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
PS Example 1; Page 24; 35pp; English.
XX
CC This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing

```

cc protein or polynucleotide sequences. The lung cancer associated

Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
sequences.

Query Match		54.3%;	Score 853.8;	DB 21;	Length 1400;
Best Local Similarity		93.3%;	Pred. No. 6,1e-196;		
Matches	934;	Conservative	1;	Mismatches	13;
				Indels	53;
				Gaps	2;
QY	607	ataaatcgaatgbybgaagaagaacgcttcttctatgcygaacaatcatgtgcatcggaaacag	666		
Db	177	agaaatcgaaatgygagaaagaacgcttcttctctatgcygaacaatcatgtgcatcggaaacag	236		
QY	667	aaactgaataagaactctgtctccaaacacgtggtgagagaaggtgcatccagtttctctat	726		
Db	237	aaccygaataagaaaactctgtctccaaacacgtggtgagagaaggtgcatccagtttctctat	296		
QY	727	gtctcgaaaacactacgtgtgactttaaaccgtgagtcagaaaccaaagaatgaaagcagctgtcta	786		
Db	297	gtctcgaaaacactacgtgtgactttaaaccgtgagtcagaaaccaaagaatgaaagcagctgtcta	356		
QY	787	gtttctctgagaagaataatcaaacagatataaagcatatcagcatgtcatctactctcc	846		
Db	357	gtttctctgagaagaataatcaaacagatataaagcatatcagcatgtcatctactctcc	416		
QY	847	agcatatagctcttcacatctcctaatacagaaagtaaaagcaaaagcactgtggaactgt	906		
Db	417	agcatatagctcttcacatctcctaatacagaaagtaaaagcaaaagcactgtggaactgt	476		
QY	907	ctctatgtagcaagtgagagcaggttcgtctatctgacaaaactaagtaaaaaataccaggtata	966		
Db	477	ctctatgtagcaagtgagagcaggttcgtctatctgacaaaactaagtaaaaaataccaggtata	536		
QY	967	caaatgagcaatgagctcagaaaacctataatacctagctccctcgaaagtgaggaacgtttgatat	1026		
Db	537	caaatgagcaatgagctcagaaaacctataatacctagctccctcgaaagtgaggaacgtttgatat	596		
QY	1027	atgatttgggacacaaatatctgt-----	1050		
Db	597	atgatttgggacacaaatatctgttttaacaattgaaacttcgagatacgggcacatacagat	656		
QY	1051	-----ttaactcaaaacccaactgtgaagaagcttttgcgtgctctcta	1094		
Db	657	tctgtgctcggagagcgttctacacaaacccaactgtgaagaagcttttgcgtgctctcta	716		
QY	1095	aaataagcttggcagtgcatatagaaatgtcttaatgccccatgatttatcatctgtctccg	1154		
Db	717	aaataagcttggcagtgcatatagaaatgtcttaatgccccatgatttatcatctgtctccg	776		
QY	1155	tattttaattacacgaatcccgacgaagaccaaatacatgttatcagaattattttaagtttt	1214		
Db	777	tattttaattacacgaatcccgacgaagaccaaatacatgttatcagaattattttaagtttt	836		
QY	1215	atccgttaatttgaataaagaattttcccatccctgtgcttcgttcagagaagaacttaataagt	1274		
Db	837	atccgttaatttgaataaagaattttcccatccctgtgcttcgttcagagaagaacttaataagt	896		
QY	1275	gcatacttggcattaaagacagactagggltcaatgctctttttaacccctttaaaaaaaatgt	1334		
Db	897	gcatacttggcattaaagacagactagggltcaatgctctttttaacccctt-aaaaaaaatgt	955		
QY	1335	taaaaggtctatgttcctcctcttcttgatcttctgaaagtttgaagtttgaactaagcatcttaagca	1394		
Db	956	taaaaggtctatgttcctcctcttcttgatcttctgaaagtttgaactaagcatcttaagca	1015		

QY	1395	aacttcgcgcttttaccagccatctcaagaatttaacaaagttcatctcaagctat	1454
Db	1016	aacttcgcgcttttaccagccatctcaagaagttaatcaaaagttcatctcaagctat	1075
QY	1455	catggatgccacacccaacaaagaagggggtgcacagaatataatcaaaagatttcgcctc	1514
Db	1076	catggatgccacacccaacaaagaaggggtgcacagaatataatcaaaagatttcgcctc	1135
QY	1515	aaattttcaataaattctctctctctccttttaaaaaaaaaa	1555
Db	1136	aaattttcaataaattctctctctctctcctttttgaaatactaa	1176

RESULT 7
 AAH85957 standard; DNA; 333 BP.
 ID AAH85957 standard; DNA; 333 BP.
 XX
 AC AAH85957;
 DT 27-FEB-2002 (first entry)
 XX
 DE Human single nucleotide polymorphism containing DNA sequence #814.
 XX
 KW Biallelic marker; polymorphism; human; disease; diagnosis; treatment;
 KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;
 KW transgenic; single nucleotide polymorphism; SNP; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Variation replace(274,C,G)
 FT /*tag_a
 XX /standard_name="single nucleotide polymorphism"
 XX
 PN WO953095-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 30-MAR-1999; 99WO-US06893.
 XX
 PR 09-APR-1998; 98US-005787L.
 XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX
 PI Lander ES, Wang D, Hudson T;
 XX
 DR WPI; 1999-620443/53.
 XX
 PT Polymorphic human genomic sequences and related allele-specific probes
 PT and primers, useful for genetic analysis, e.g. diagnosis and monitoring
 PT of disease -
 XX
 PS Claim 1; Page 117; 330pp; English.
 XX
 CC This invention describes novel human nucleic acid segments (I)
 CC containing polymorphic sites. The polynucleotides of (I) are used for,
 CC e.g. correlating disease polymorphisms (or disease susceptibility) or
 CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,
 CC response to drugs etc.); diagnosing and monitoring e.g. cancer,
 CC inflammation, heart or central nervous system diseases; detecting
 CC susceptibility to microbial infection; treating or preventing such
 CC diseases; forensic analysis; gene therapy; paternity testing; mapping
 CC genomic loci associated with phenotypic traits (and subsequent cloning
 CC of the genes responsible); and the production of transgenic organisms.
 CC Antibodies raised against (I) are useful as diagnostic and therapeutic
 CC tools and in drug screening. AAH85144 - AAH87644 represent the human
 CC DNA sequences containing biallelic polymorphic sites described in the
 CC invention.
 XX
 SQ Sequence 333 BP; 90 A; 64 C; 53 G; 117 T; 9 other:

Query Match 20.6% Score 324; DB 20; Length 333;
 Best Local Similarity 97.3% Pred. No. 1.7e-68;
 Matches 324; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1159 ttaatttactgattccagcaagaacccaatcattgatacagattattttaagtttatcc 1218
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ttaatttactgattccagcaagaacccaatcattgatacagattattttaagtttatcc 60

QY 1219 gtaatttgataaaaagaatttccattccttggtctgtcagagaacctataaagtcta 1278
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 gtagtttgataaaaagaatttccattccttggtctgtcagagaacctataaagtcta 120

QY 1279 ctgtgcattagagcagactagggttcattctctttaccctttaaaaaaatgttaa 1338
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 ctgtgcattagagcagactagggttcattctctttaccctttaaaaaaatgttaa 180

QY 1339 agtctaagttactactttttcttatttgcagcttgcagcttgcagcattcagaacct 1398
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 agtctaagttactactttttcttatttgcagcttgcagcttgcagcattcagaacct 240

QY 1399 tcgaagtttgactagcactcctcaagaagtttaatacaagaatcattcagctgtacatt 1458
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 tcgaagtttgactagcactcctcaagaagtttaatacaagaatcattcagctgtacatt 300

QY 1459 ggatcctactcaacaaaaggagggtgtcaga 1491
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 ggatcctactcaacaaaaggagggtgtcaga 333

RESULT 8
 AAH87390
 ID AAH87390 standard; DNA; 333 BP.
 AC
 AAH87390;
 XX
 27-FEB-2002 (first entry)
 DE Human single nucleotide polymorphism containing DNA sequence #2247.
 DE
 XX Biallelic marker; polymorphism; human; disease; diagnosis; treatment;
 KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;
 KW transgenic; single nucleotide polymorphism; SNP; ss.
 OS Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT replace(238,6)
 FT /**tag=
 FT /standard_name="single nucleotide polymorphism"
 FT
 XX
 PN WO953095-A2.
 PN
 XX
 PD 21-OCT-1999.
 PD
 XX
 PF 30-MAR-1999; 99WO-US06893.
 PF
 XX
 PR 09-APR-1998; 98US-0057871.
 PR
 XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA
 XX
 PI Lander ES, Wang D, Hudson T;
 PI
 XX
 DR WPI; 1999-620443/53.
 DR
 XX
 PT Polymorphic human genomic sequences and related allele-specific probes
 PT and primers, useful for genetic analysis, e.g. diagnosis and monitoring
 PT of disease -
 PT
 XX
 PS Claim 1; Page 283; 330pp; English.
 PS
 XX
 CC This invention describes novel human nucleotide acid segments (I)
 CC containing polymorphic sites. The polymorphisms of (I) are used for,
 CC e.g. correlating disease polymorphisms (or disease susceptibility) or

CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,
 CC response to drugs etc.); diagnosing and monitoring e.g. cancer,
 CC inflammation, heart or central nervous system diseases; detecting
 CC susceptibility to microbial infection; treating or preventing such
 CC diseases; forensic analysis; gene therapy; paternity testing; mapping
 CC genomic loci associated with phenotypic traits (and subsequent cloning
 CC of the genes responsible); and the production of transgenic organisms.
 CC Antibodies raised against (I) are useful as diagnostic and therapeutic
 CC tools and in drug screening. AAH85144 - AAH87644 represent the human
 CC DNA sequences containing biallelic polymorphic sites described in the
 CC invention.
 XX
 SQ Sequence 333 BP; 90 A; 64 C; 53 G; 117 T; 9 other;

Query Match 20.6% Score 324; DB 20; Length 333;
 Best Local Similarity 97.3% Pred. No. 1.7e-68;
 Matches 324; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1159 ttaatttactgattccagcaagaacccaatcattgatacagattattttaagtttatcc 1218
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 ttaatttactgattccagcaagaacccaatcattgatacagattattttaagtttatcc 60

QY 1219 gtaatttgataaaaagaatttccattccttggtctgtcagagaacctataaagtcta 1278
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 gtagtttgataaaaagaatttccattccttggtctgtcagagaacctataaagtcta 120

QY 1279 ctgtgcattagagcagactagggttcattctctttaccctttaaaaaaatgttaa 1338
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 ctgtgcattagagcagactagggttcattctctttaccctttaaaaaaatgttaa 180

QY 1339 agtctaagttactactttttcttatttgcagcttgcagcttgcagcattcagaacct 1398
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 agtctaagttactactttttcttatttgcagcttgcagcttgcagcattcagaacct 240

QY 1399 tcgaagtttgactagcactcctcaagaagtttaatacaagaatcattcagctgtacatt 1458
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 tcgaagtttgactagcactcctcaagaagtttaatacaagaatcattcagctgtacatt 300

QY 1459 ggatcctactcaacaaaaggagggtgtcaga 1491
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 ggatcctactcaacaaaaggagggtgtcaga 333

RESULT 9
 AAX11374
 ID AAX11374 standard; DNA; 251 BP.
 AC
 AAX11374;
 XX
 30-MAR-1999 (first entry)
 DE Human biallelic polymorphic DNA fragment WI-7773b.
 DE
 XX
 DE
 XX Biallelic marker; polymorphism; human; forensic; paternity testing; disease;
 KW polymorphism; biallelic; human; forensic; paternity testing; disease;
 KW detection; phenotypic typing; characteristic; infection; hereditary;
 KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
 KW treatment; marker; ss.
 OS Homo sapiens.
 OS
 XX
 PN WO9820165-A2.
 PN
 XX
 PD 14-MAY-1998.
 PD
 XX
 PF 05-NOV-1997; 97WO-US20313.
 PF
 XX
 PR 06-NOV-1996; 96US-0030455.
 PR
 XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA
 XX
 PI Hudson T, Lander ES, Wang D;
 PI
 XX

DR WPI: 1998-286974/25.

PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease

PS Claim 1; page 152; 310pp; English.

AX10265X12937 are human DNA fragments which contain biallelic polymorphic markers which have been isolated using the primers represented in AXA09121-X10268. The base occupying the polymorphic site is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments can be used in methods for determining polymorphic forms in an individual for use in e.g. forensics, paternity testing or for phenotypic typing for diseases such as agammaglobulinemia, diabetes insipidus, Leech-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, polycystic kidney disease, hereditary spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, acute intermittent porphyria, autoimmune diseases, inflammatory, cancer, diseases of the nervous system, infection by pathogenic microorganisms, and characteristics such as longevity, appearance (e.g. baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments. The isolated polymorphic nucleic acid segments can also be used to produce medicaments for the treatment or prophylaxis of such diseases.

Sequence 251 BP; 62 A; 44 C; 37 G; 98 T; 10 other;

Query Match	15.4%	Score 241.6;	DB 19;	Length 251;
Best Local Similarity	96.0%;	Pred. No. 1.1e-48;		
Matches 241; Conservative	1;	Mismatches 9;	Indels 0;	Gaps 0;

Qy 1159 ttaattactgattccagcaagaccacaatctgtatcagattattttaagttatcc 1218
|||||
Ddb 1 ttaattactgattccagcaagaccacaatctgtatcagattattttaagttatcc 60

key	Location/Qualifiers
primer_bind	complement (1..50)
primer_bind	/tag= a

```

FT /note= "Mature CPB 5'-end primer"
FT 898..927
FT primer_bind
FT /tag= b
FT /note= "CPB 3' end primer"

```

PN WO9623064-A1

PD 01-AUG-1996.

PF 25-JAN-1996; 96WO-US00995.

PR 25-JAN-1995; 95US-0378233.

PA (BIOT-) BIOTECHNOLOGY GEN CORP.

PI Fulga N, Gorecki M, Hartman J, Mendelovitch S,

DR WPI; 1996-362688/36.

DR P-PSDB; AAW00602.

PT Purified active recombinant carboxypeptidase B prodn. - by
PT expressing DNA encoding the pro-enzyme, folding and enzymatic
PT cleavage to give active form, used e.g. for insulin prodn.

PS Example 1; Page 37-38; 49pp; English

A cDNA sequence (AA135760) codes for the mature form (AA000602) of rat carboxypeptidase B (CPB). A full-length sequence including cDNA (AA135759) coding for the activation peptide (AA000601) can be obtained by PCR amplification (see also AA135758 and AA135758) of cDNA derived from Sprague-Dawley rat pancreas. Pro-CPB can be expressed in host cells, e.g. *Escherichia coli*, and subsequently recovered, CPB refolded and cleaved with trypsin to yield the active enzyme. CPB produced this way is cheaper than porcine pancreatic enzyme, and is free of other proteases.

Sequence 927 BP; 253 A; 226 C; 213 G; 235 T; 0 other;

Query Match	7.9%	Score 124;	DB 17;	Length 927;
Best Local Similarity	57.9%	Pred. No. 3.6e-20;		
Matches 278; Conservative	0;	Mismatches 195;	Indels 7;	Gaps 3;

Matches	278;	Conservative	0;	Mismatches	195;	Indels	7;	Gaps
Matches	278;	Conservative	0;	Mismatches	195;	Indels	7;	Gaps


```

FT      Variation                                replace(88,C)
FT      FT                                     /*tag=
FT      XX                                     a
XX      PN      MO9953095-A2.                  /standard_name= "single nucleotide polymorphism"
XX      PD      21-OCT-1999.
XX      PE      30-MAR-1999;    99WO-US06893.
XX      PR      09-APR-1998;    98US-0057871.
XX      PA      (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX      PI      Lander ES, Wang D, Hudson T;
XX      DR      WPI: 1999-620443/53.
XX      PT      Polymorphic human genomic sequences and related allele-specific probes
XX      PR      and primers, useful for genetic analysis, e.g. diagnosis and monitoring
XX      PS      Claim 1; Page 91; 330pp; English.
XX      CC      This invention describes novel human nucleic acid segments (I)
CC      containing polymorphic sites. The polynucleotides of (I) are used for,
CC      e.g. correlating disease polymorphisms (or disease susceptibility) or
CC      other phenotypic traits (e.g. baldness, obesity, fertility, strength,
CC      response to drugs etc.); diagnosing and monitoring e.g. cancer,
CC      inflammation, heart or central nervous system diseases; detecting
CC      susceptibility to microbial infection; treating or preventing such
CC      diseases; forensic analysis; gene therapy; paternity testing; mapping
CC      genomic loci associated with phenotypic traits (and subsequent cloning
CC      of the genes responsible); and the production of transgenic organisms.
CC      Antibodies raised against (I) are useful as diagnostic and therapeutic
CC      tools and in drug screening. AAH85144 - AAH87644 represent the human
CC      DNA sequences containing biallelic polymorphic sites described in the
CC      invention.
SQ      Sequence 121 BP; 33 A; 22 C; 29 G; 36 T; 1 other;

Query Match          7.6%; Score 120; DB 20; Length 121;
Best Local Similarity 99.2%; Pred. No. 1.7e-19;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

OY      1363 gattttgacgttactcagcatctcaagaaccttttgacgtttgactcagcatctcaa 1422
        |||
DB      121 GAATTTTCGACGTTTTACACAGCCATCTCAAGCAACTTTCGACGTTCACATCTTCAA 62
OY      1423 gaagttaataacaagatcatctcacgcgtgatcatgtgaacctactcaacaagaagaa 1482
        |||
DB      61 GCAAAGTTTAATCAAGAATCATCTCACGCGTGATCATTGGTCTCTACTACACAAAAGGAAG 2
OY      1483 g 1483
        |
DB      1 G 1

RESULT 15
AAVA1795
ID      AAVA1795 standard; DNA; 1263 BP.
XX      AC      AAVA1795;
XX      DT      20-NOV-1998 (first entry)
XX      DE      Human pancreatic carboxypeptidase nucleotide sequence.
XX      KW      ss: human; pancreatic carboxypeptidase B; insulin; protein sequencing;
XX      KM      prodruq therapy.
XX      OS      Homo sapiens.
```

[illegible]

Db 1002 agaacctggcc---tcactgcacgacccaagtlacacatalgtgccggaagctacacacat 1058
QY 992 ataccctgctcctggaggtggggaagattggaatgatgattgggaatcaataatcgtt 1051
Db 1059 ctaccctgctcctgggagctctgaagactggcttatgacaaaggaatcagatatcctt 1118
QY 1052 tacat 1056
Db 1119 cacct 1123

Search completed: September 18, 2002, 16:35:58
Job time: 4316 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2002, 15:25:32 ; Search time 54.5 Seconds

(without alignments)
7089.569 Million cell updates/sec

Title: US-09-980-881-1

Perfect score: 1573

Sequence: 1 agaaatgctgttggatg.....aaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCrUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1372	87.2	1749	1	US-07-649-591B-2
2	1372	87.2	1749	1	US-08-277-540-2
3	1372	87.2	1749	1	US-08-430-787A-2
4	918	58.4	1272	2	US-08-869-057-1
5	124	7.9	927	2	US-08-782-760-5
6	124	7.9	927	5	PCr-US96-00995-5
7	123.4	7.8	921	1	US-08-696-139-3
8	123.4	7.8	1215	1	US-08-696-139-1
9	107.4	6.8	999	2	US-08-860-882A-67
10	107.4	6.8	1053	2	US-08-860-882A-64
11	107.4	6.8	1263	2	US-08-860-882A-56
12	107.4	6.8	1284	2	US-08-860-882A-71
13	104.2	6.6	1059	2	US-08-860-882A-74
14	104.2	6.6	1059	2	US-08-860-882A-77
15	101	6.4	1870	4	US-09-171-945-112
16	101	6.4	2154	4	US-09-171-945-124
17	71.6	4.6	1251	4	US-08-640-906-3
18	70.6	4.5	1251	4	US-09-395-936-3
19	70.6	4.5	1257	4	US-08-640-906-1
20	70.6	4.5	1257	4	US-09-395-936-1
21	46.8	3.0	285	2	US-08-782-760-2
22	46.8	3.0	285	5	PCr-US96-00995-2
23	45.6	2.9	8920	4	US-08-446-855A-1
24	45.6	2.9	8920	4	US-09-150-741-1
25	45	2.9	4665	4	US-08-948-378A-7
26	45	2.9	4665	4	US-09-169-425C-7
27	44.8	2.8	7218	1	US-08-232-463-14

28	43.8	2.8	880	1	US-08-616-368A-7	Sequence 7, Appl
29	43.8	2.8	880	3	US-09-054-298-7	Sequence 7, Appl
30	43.8	2.8	880	4	US-08-818-655-7	Sequence 7, Appl
31	43.8	2.8	1610	4	US-09-013-896A-1	Sequence 10, Appl
32	43.2	2.7	9521	4	US-08-972-218-2	Sequence 2, Appl
33	43.2	2.7	13905	4	US-08-872-218-1	Sequence 1, Appl
34	42.4	2.7	1129	4	US-09-227-357-40	Sequence 40, Appl
35	42.4	2.7	1472	4	US-08-781-420-10	Sequence 10, Appl
36	42.4	2.7	1472	4	US-08-781-420-12	Sequence 12, Appl
37	42.4	2.7	1472	4	US-08-874-102-10	Sequence 10, Appl
38	42.4	2.7	1472	4	US-08-874-102-12	Sequence 12, Appl
39	42.4	2.7	1881	4	US-08-874-102-46	Sequence 46, Appl
40	42.4	2.7	1881	4	US-08-874-102-48	Sequence 48, Appl
41	42	2.7	2158	1	US-08-698-551-1	Sequence 1, Appl
42	42	2.7	2158	1	US-08-602-228-1	Sequence 1, Appl
43	42	2.7	2158	2	US-08-649-341A-1	Sequence 1, Appl
44	42	2.7	2158	2	US-08-494-440B-1	Sequence 1, Appl
45	42	2.7	2158	2	US-08-533-901B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Dryna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:

OTHER INFORMATION:
 NAME/KEY: signal sequence
 LOCATION: 41 to 106
 IDENTIFICATION METHOD:
 OTHER INFORMATION:
 US-07-649-591B-2

Query Match 87.2%; Score 1372; DB 1; Length 1749;
 Best Local Similarity 90.3%; Pred. No. 0;
 Matches 1558; Conservative 0; Mismatches 5; Indels 165; Gaps 2;

1 agaaaaattgcttggatgaagcttgcagccttgacgtccttgtaaccattgtctct 60
 24 aaaaaattgcttggatgaagcttgcagccttgacgtccttgtaaccattgtctct 83
 61 tcttgagagagatctctgcgcttcagagcttgagagcttgaagcttgcagccttg 120
 84 tcttgagagagatctctgcgcttcagagcttgagagcttgaagcttgcagccttg 143
 121 cctctagagagcttgcagccttgacgtccttgtaaccattgtctctgc 180
 144 cctctagagagcttgcagccttgacgtccttgtaaccattgtctctgc 203
 181 agcggtaacagctgacgtccttgtaaccattgtctctgc 240
 204 agcggtaacagctgacgtccttgtaaccattgtctctgc 263
 241 ctgagctgagagcttgcagccttgacgtccttgtaaccattgtctctgc 300
 264 ctgagctgagagcttgcagccttgacgtccttgtaaccattgtctctgc 323
 301 tggagagctgagagcttgcagccttgacgtccttgtaaccattgtctctgc 360
 324 tggagagctgagagcttgcagccttgacgtccttgtaaccattgtctctgc 383
 361 cctctagagagcttgcagccttgacgtccttgtaaccattgtctctgc 420
 384 cctctagagagcttgcagccttgacgtccttgtaaccattgtctctgc 443
 421 ttaatactgagagcttgcagccttgacgtccttgtaaccattgtctctgc 480
 444 ttaatactgagagcttgcagccttgacgtccttgtaaccattgtctctgc 503
 481 agtaccacactgagcttgcagccttgacgtccttgtaaccattgtctctgc 540
 504 agtaccacactgagcttgcagccttgacgtccttgtaaccattgtctctgc 563
 541 ggaatgagctgagagcttgcagccttgacgtccttgtaaccattgtctctgc 600
 564 ggaatgagctgagagcttgcagccttgacgtccttgtaaccattgtctctgc 623
 601 tggagagcttgcagccttgacgtccttgtaaccattgtctctgc 608
 624 tggagagcttgcagccttgacgtccttgtaaccattgtctctgc 683
 609 -----a 609
 684 tggatcttatgtatgacgggtgattatgacgggtgattatgacgggtgattatgac 743
 610 atcgaaatgag 669
 744 atcgaaatgag 803
 670 tgaatgag 729
 804 tgaatgag 863
 730 cggag 789
 864 cggag 923

790 tcttgag 849
 924 tcttgag 983
 850 atatagtgttccatattccttaacagagagagagagagagagagagagagagag 909
 984 atatagtgttccatattccttaacagagagagagagagagagagagagagagag 1043
 910 tagtgcag 969
 1044 tagtgcag 1103
 970 atgag 1029
 1104 atgag 1163
 1030 atttgag 1050
 1164 atttgag 1223
 1051 -----ttacatcaaacacagctgagagagagagagagagagagagagag 1097
 1224 tgcctgcag 1283
 1098 tagcttgag 1157
 1284 tagcttgag 1343
 1158 ttaatactgag 1217
 1344 ttaatactgag 1403
 1218 cgtagcttgag 1277
 1404 cgtagcttgag 1463
 1278 acttgag 1337
 1464 acttgag 1523
 1338 aagcttgag 1397
 1524 aagcttgag 1583
 1398 ttcgagcttgag 1457
 1584 ttcgagcttgag 1643
 1458 tggag 1517
 1644 tggag 1703
 1518 tttcaataatattct 1563
 1704 tttcaataatattct 1749

RESULT 2
 US-08-277-540-2
 Sequence 2, Application US/08277540
 Patent No. 5474901
 GENERAL INFORMATION:
 APPLICANT: Drayna, Dennis T., Eaton, Dan L.
 TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:

QY 1458 tggatcctactcaacaaaggaaggtgycagagatcaataaagattctgtcccaaa 1517
 Db 1644 TGGATCCTACTCAACAAAGGAGGCTGTCAGAGTACATTAAAGATTCTGCTCCAAA 1703
 QY 1518 ttccaataattct 1563
 Db 1704 TTTTCATTAATTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1749

RESULT 3
 US-08-430-787A-2
 ; Sequence 2, Application US/08430787A
 ; Patent No. 5593674
 ; GENERAL INFORMATION:
 ; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
 ; TITLE OF INVENTION: NO. 5593674el Plasma Carboxypeptidase
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 460 Point San Bruno Blvd
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080

COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: patin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/430,787A
 FILING DATE: 27-APR-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/277,540
 FILING DATE: 19-JUL-1994
 APPLICATION NUMBER: 08/167727
 FILING DATE: 15-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/959944
 FILING DATE: 14-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/649591
 FILING DATE: 01-FEB-91
 ATTORNEY/AGENT INFORMATION:
 NAME: Hasak, Janet E.
 REGISTRATION NUMBER: 28,616
 REFERENCE/DOCKET NUMBER: 689D1C1D1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-1896
 TELEFAX: 415/952-9881
 TELE: 910/371-7168
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1749 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-430-787A-2

Query Match 87.2%; Score 1372; DB 1; Length 1749;
 Best Local Similarity 90.3%; Pred. No. 0;
 Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

QY 1 agaaatgctgttggaatgaagcttgcagccttgcaagtcctgttaaccaatgttctct 60
 Db 24 AGAAATGCTGTGGAGTGAAGCTTGCAGCTTGCAGTCTGTGACCATGTTCTCTCT 83
 QY 61 tctgtgagcagcatgtcttcgcttcagagtggtgcagagttctagctgtctctctctaa 120
 Db 84 TCTGTGACGACATGTTCTTCTGCGTTTCAAGTGGCAAGTTTCAAGTCTCTCTCTGAA 143

QY 121 cctctgagcaagtccaagtctctacagaatctctactacaacatatgagattgtctctg 180
 Db 144 CCTCTAGGCAAGTTCAGATTCTTACAGAACTTACTACAAACATATGAGATTGTCCTG 203
 QY 181 agccggttaacagcttgacccctattgtgaagaaaaaagaagtcatttttgaatagct 240
 Db 204 AGCCGGTAACAGCTGACCTTATTGTGAACAAAACAACTCCATTTTGTGTAATGCT 263
 QY 241 ctgagtgcacaatgtgaagcccatctaaatgtgagcggaaattccatgagtgctgc 300
 Db 264 CTGATGTCACATGTGAAGCCCATTTAAATGTGAGCGAATTCATGCACTGCTTGC 323
 QY 301 tggcagagctggaaagctctattcaacagcaatttccaaagcagtcagcccgag 360
 Db 324 TGGCAGACCTGGAAGTCTTATTCAACAGCATTTCCAAACACACAGTCCACCCCGAG 383
 QY 361 cctccgcatcgtactatgacagatgcactcactaaatgaatctctcttgatagat 420
 Db 384 CCTCCGATCGTACTATGAGATGACATCTCACTCAATGAATCTTATCTTGATGAT 443
 QY 421 ttaactgagagggacatctctgatactgcttaacaaaatccaatggatccctattg 480
 Db 444 TTATACCTGAGAGCATCTCTGATATGCTTACAAAATCCATTTGATGATCTTGA 503
 QY 481 agtaccacatctatgtttaaggttcttggaagaaagcaacagcccaaaatgcatat 540
 Db 504 ACTACCACTCATTTTAAAGTTTCTGCAAAAAGAACAAACAGCCAAAATGCTCATAT 563
 QY 541 ggaattgactgtgaatccatgcagagaatgatatctctctctctctgtgtgtca 600
 Db 564 GGATTGACTGTGAATCCATGCGCAGAGATGATCTCTGCTTGTGTTGGTTCA 623
 QY 601 taagccat----- 608
 Db 624 TAGGCATATACCTCAATTTCTATGGGAATTAAGGCAATATACCAATCTCTGAGCTTG 683
 QY 609 -----a 609
 Db 684 TGGATTCTATGTTATGCGGGTGTATGTGAGAGGTTATGACTACTCATGAAAAAGA 743
 QY 610 atggaatgtggaagaaagacgcttctctctctctctctctctctctctctctctct 669
 Db 744 ATCGAATGTGAAGAAAGACGTTCTTCTATGGAACAACTATGATGCGAAGACAGC 803
 QY 670 tgaatgaacattgtctcctcaacacatgtgttggaagatgcatcaagtctctctgct 729
 Db 804 TGAATGAGAACTTTGCTTCCAAACACATGCTGTGAGAAAGTGCAGATCCCATGCT 863
 QY 730 cggaaacctactgtgacattatctctgagtcagaaacagaaagtgaaagcagtgctag 789
 Db 864 CGGAACCTACTGTGACTTATCTGAGTCAGAACCAAGAAAGTGAAGCAGTGGCTAGTT 923
 QY 790 tcttgagaagaataatacaacagattaaagatatacaatgacatgcatctactccagc 849
 Db 924 TCTTGAAGAAATATACACAGATTAAGCATATACACATGATCTTACTCTCCAGC 983
 QY 850 atatagtgcttccatctctctatacagaaatgaagtaaaagcaagacatgaaagctctc 909
 Db 984 ATATAGTGTTCATATTCATATTCATACAGAACTAAACCAAGCAATGAGAACTGTCTC 1043
 QY 910 taagtagcagtgaaagcagctcgctgctatgtgacaanaactagtaaaataccaggtata 969
 Db 1044 TAGTAGCCATGAGAGCGTTCGTGTATGAGAAACTAGTAAATAATACAGATATACAC 1103
 QY 970 atgagcattgctcagaacatttaccatgctctctctctctctctctctctctctct 1029
 Db 1104 ATGGCCATGCTCAGAAACCTTATACCTTACCTCTGAGTGGAGCATGATGATG 1163
 QY 1030 attgggcatcaaatatcgt----- 1050
 Db 1164 ATTGGGCAATCAATATTCGTTTACATTTGAATTCGAGATACGGGACATACGATTTCT 1223
 QY 1051 -----ttacatcaaacccaactgttagagaagcttctcgctgtctctctctctct 1097

```

Db 1224 TCGTCGCGGAGGCTTACACAAACCACCTGAGAGAACCTTTGGCGTCTCTAATAA 1283
Oy 1098 tagcttgcatgtcattagaaatgttaatgccccgattatcatctgtcttcgatat 1157
Db 1284 TAGCTTGCGATGTCTATAGGAATGTTAATGCCCCGATTATATCTTGTGCTCCGAT 1343
Oy 1158 tttaattactgattccagaacaaacalcattatcatatttttaagtttatac 1217
Db 1344 TTTAATTTACTGATTCACACAGACCAATATGATGATGATTAATTTTAAAGTTTATC 1403
Oy 1218 cgtatcttgataaaagatttccattccttggtgtctgtcagaagaactataatgtct 1277
Db 1404 CGTACTTTTGATTAAGAATTTTCATTCCTTGTTCTGACAGAACTAATTAATGTCT 1463
Oy 1278 acttgccattaagcagacagaggttcacgtcttttacccttaaaaaaatgttaa 1337
Db 1464 ACTTGGCCATTAAAGCAGACTAGGGTCTATGCTTTTTCACCTTTAAAAAAATTTGTA 1523
Oy 1338 aagctagttactactcttcttcttgatttcgaagctttgactagcactcacaagaact 1397
Db 1524 AAGCTAGTACTACTACTTTCTTTGATTTTTCGACCTTGACTAGCCATCTCAAGCACT 1583
Oy 1398 ttccagcgtttgactagccatctccagaagaatttaataagaatcatctcaagctgatac 1457
Db 1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAAGATCATCTACGCTGATCAT 1643
Oy 1458 tggactcactcaacaaagaagagtggtcagaagttacattaaagatttcgtccaaa 1517
Db 1644 TGGATCTCTCTACAAAAGGAAGGTTGTCAGAAAGTACATTTAAGATTTCGCTCAAA 1703
Oy 1518 tttaataaattcttcttcttcttcttaaaaaaaaaaaaaaaaaa 1563
Db 1704 TTTTCAATTAATTTCTTCTCTCTTAATAAAAAAAAAAAAAA 1749

RESULT 4
US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Mariko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94804-0099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/869, 057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: A35
; ATTORNEY/AGENT INFORMATION:
; NAME: Washlien, Wendy L
; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-262-3411
; TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1272 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PUBLICATION INFORMATION:
; AUTHORS: Eaton, Dan L.
; AUTHORS: Malloy, Beth E.
; AUTHORS: Tsai, Siao P
; AUTHORS: Hensel, William
; AUTHORS: Drayna, Dennis
; TITLE: Isolation, Molecular Cloning, and Partial
; TITLE: Characterization of a No. 5985562el Carboxypeptidase B
; JOURNAL: J. Biol. Chem.
; VOLUME: 266
; ISSUE: 32
; PAGES: 21833-21838
; DATE: No. 5985562 15-1991
; US-08-869-057-1

Query Match      58.4%; Score 918; DB 2; Length 1272;
Best Local Similarity 86.8%; Pred. No. 7.2e-213;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

Oy 18 atgaagcttgcagccttcgacgtcccttgtaacccattgtctctctgtgagcagcatgctc 77
Db 1 ATGAAGCTTTGCAGCCTTGCAGCTTGTGACCTATGTTCTCTTGTGAGCAGCATGTC 60
Oy 78 ttgcggttccagagtggtgccaagtcttagctgtcttccctagaaccccttggaagttcaa 137
Db 61 TTGCGCTTTCAGAGTGCGCCAAAGTTCTAGCTGCTCTCTTAAGACCTTAGCGAAGTTCAA 120
Oy 138 gtctcaagactcttaactcaacatatgagattgtctctctgcagccggtgaacagctgac 197
Db 121 GTCTCAAGACTCTTACTCAACATATGAGATTGTTCTCGCAGCCGGTAACAGCTGAC 180
Oy 198 ctattgtgaagaagaacaagtcacatttcttgtaaatgcatctgagtgcacaatgctg 257
Db 181 CTATTGTGTAAGAAAAACAAGTCCATTGTTTGTAAATGCATGTCATCGCAATGTG 240
Oy 258 aaagccatttaattgtgaggaattccatgcaatgagtgctgtgtgcaagcgtggaagat 317
Db 241 AAAGCCATTTAATGTGAGCGGAATTCATGCAAGTGTCTGTGGCAGCGTGAAGAT 300
Oy 318 ctattcaacagcagatttccacgacacagtcagcccccagccctccgcatctactat 377
Db 301 CTATTCAACAGCAGATTTCACAGACAGACGTAGCCCCCGACCTCCGATGTAAT 360
Oy 378 gaacaglatcactcactaaatgaatctatcttcttgatagaattatlaactgagaagcat 437
Db 361 GAACAGTATCCTCACTAAATGAATCTATTTGATAGAAATTTAATCACTGAGAGGCAT 420
Oy 438 cctgatattgttcaaaaatccacattgtgactcctcatcttgagaagtaaccacttatgtt 497
Db 421 CCGTATATGCTTACAAAACCCACATTGATCCTCATTTGAGAGTACCCACTTATGT 480
Oy 498 ttaaaagttctgaaagaacacaacagccaanaaatgcatatgattgtgactgtggaac 557
Db 481 TTTAAGGTTTCTGAAAAAGAACAAACAGCCAAAATGCCATATGATGACTGTGGAATC 540
Oy 558 catgcagaagaatgagatctctctgtcttctgtgtgtgtatagagcat----- 608
Db 541 CATGCCAGAAATGATGATCTCTCTGCTTGTGCTTGTGTTGATAGGCCATATTAACATCA 600
Oy 609 ----- 608
Db 601 TTTATAGGATAATAGGGCAATATACCAATCTCTGAGCTTGATTTCTATGTTATG 660
Oy 609 -----atcgaatgtgagaag 626
Db 661 CCGGTGTTAATGTGACGCTTATGACTACTCATGAAAAAAGATGCAATGATGAGAGAA 720
Oy 627 aaccgttcttctatgcgaacaatcattgcatacgaacagacctgaaatagcaactgtgc 686

```


Db 618 TGATTACAACCTCCCGAGACATCTGATGTAATTAACCTGGCTAAGGCTGCCGTGA 677
Qy 932 tgcatttgcacaaactagtaaaaaataccaggtatcacatgycgcatgctcagaacctt 991
Db 678 AGAAGCTTG---CTACACTGTATGGCACCAAGTACACATACGGCCGAGAGCTACACACAT 734
Qy 992 ataccagtcctcggaggaggaggagagatgcatctatgatttgggcataaattatcgctt 1051
Db 735 CTATCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGAAATATTCTCTT 794
Qy 1052 tacat 1056
Db 795 CACCT 799

RESULT 8
US-08-696-139-1
; Sequence 1, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hersiberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Hailiao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8661
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1215 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1215
; US-08-696-139-1

Query Match 7.8%; Score 123.4; DB 1; Length 1215;
Best Local Similarity 56.3%; Pred. No. 5.5e-21;
Matches 273; Conservative 0; Mismatches 206; Indels 6; Gaps 2;
Qy 572 gatctctctgcttctgcttgggtcatalagccataatcgaatgtggaagaagacgcg 631

Db 615 GCTCATATTGATGGCTACATCTACACCTGAGACCAAGAACCAATGTGAGAAACACCCG 674
Qy 632 tctcttctctgcgaacaatcattgcacgcgaagacctgtaataagcaattgtctccaa 691
Db 675 CTCTACCAATGCTGGAAACTACCTGCATTGGCACAGACCCCAACAGAAATTTGATGCTGG 734
Qy 692 acacgtgtgtggaaggtgcattccagttccatgcctgcggaacctctgtggacttta 751
Db 735 G---TGTCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGC 791
Qy 752 tccgtgaatgaagaacgaagtgagcgagtgcttcttctgagaagaatlatcaacca 811
Db 792 TCCAGAGCTGTGAATAAAGAGACCAAGCCCTGGCTGATTATTACCAACCAACCTCTCTC 851
Qy 812 gattaaagcatcacatcagcatgcatctatctatccagcatatagtgttccatctcta 871
Db 853 CATCAAGCATACCTGACGATCCACTCATACACAGATGATACTTACCTTATTCCTA 911
Qy 872 tacacgaagtaaaagcaaaagccatgaggaactgtctctagtagccagtgagaagttcg 931
Db 912 TGATTACAACCTCCCGAGAACATGCTGATGTAATACCTGGCTAAGGCTGCCGTGA 971
Qy 932 tgcatttgcacaaactagtaaaaaataccaggtatcacatgycgcatgctcagaacctt 991
Db 972 AGAAGCTTG---CTACACTGTATGGCACCAAGTACACATACGCGCCGAGAGCTACACACAT 1028
Qy 992 ataccagtcctcggaggaggaggagagatgcatctatgatttgggcataaattatcgctt 1051
Db 1029 CTATCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGAAATATTCTCTT 1088
Qy 1052 tacat 1056
Db 1089 CACCT 1093

RESULT 9
US-08-860-882A-67
; Sequence 67, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGLETE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAH, JOHN FREDERICK
; APPLICANT: BLAKLEY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID HOW
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653

Db 672 GCTCAATATTGATGGCTACATCTACACCTGACCAAGAGCCGATTTTGAGAAAGACTCG 731
Qy 632 ttcttctatgagaaacatcatgacggaacagacctgaaatagcaacttgtctccaa 691
Db 732 CTCACCCCATGATGATCTAGCTGCAATGCGACAGACCCCAACAGAAATTTGATGCG--- 788
Qy 692 acactggtgtggaagagtgatccagttccatctgctggaactactatgttgacttta 751
Db 789 TGGTGGTGTGAAATTTGGAGCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 848
Qy 752 tcttgatgcaagcaagatgaaggcagtggtgcttcttcttgaagaataatacaaca 811
Db 849 CCGAGAGTGTGAAAGGACACCAAGCCCTGGCTGATTTTCATCCGACAAACTCTCTTC 908
Qy 812 gattaaagcatatcagcagatgcatctactccagcatatagtggttccatattccaa 871
Db 909 CATCAAGGATATCTGACATCCACTCGTACCTCCCAATGATGATCTACCTACTCANA 968
Qy 872 tacaggaagtaagcaagacatgaggaactgtctctagtagcagtggaagcagttcg 931
Db 969 TCTTACAACTCGGTGAGAACATGCTGATGAAATGCTGCTAAAGCTACTGTGAA 1028
Qy 932 tgcatttgcgaacaaactagtaaaataccaggtatatacacatggccatgctcagaacact 991
Db 1029 AGACTTGGC---TCACCTGACGGGACCAAGTACACATATGGCCCGGAGCTACACAT 1085
Qy 992 ataccatgctcctgaggtggtggaagcattggtatgatttggtggaatcaaatatcg 1051
Db 1086 CATCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGATGATATTCCTT 1145
Qy 1052 tacat 1056
Db 1146 CACCT 1150

RESULT 13
US-08-860-882A-74
; Sequence 74, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1059 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-860-882A-74

Query Match 6.6%; Score 104.2; DB 2; Length 1059;
Best Local Similarity 53.8%; Pred. No. 2,3e-16;
Matches 261; Conservative 0; Mismatches 218; Indels 6; Gaps 2;

Qy 572 gactctccgcttctgtctgtgtgttcataagccataactgagatgtggaagaagaccg 631
Db 387 GCTCAATATTGATGGCTACTACTACACTGACCAAGAGCCGATTTTGAGAAAGACTCG 446
Qy 632 ttcttctatgagaaacatcatgacggaacagacactgaaatagcaacttgtctccaa 691
Db 447 CTCACCCCATGATGATCTAGCTGCAATGCGACAGACCCCAACAGAAATTTGATGCG--- 503
Qy 692 acactggtgtggaagagtgatccagttccatctgctggaactactatgttgacttta 751
Db 504 TGGTGGTGTGAAATTTGGAGCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563
Qy 752 tcttgatgcaagcaagatgaaggcagtggtgcttcttcttgaagaataatacaaca 811
Db 564 CCGAGAGTGTGAAAGGACACCAAGCCCTGGCTGATTTTCATCCGACAAACTCTCTTC 623
Qy 812 gattaaagcatatcagcagatgcatctactccagcatatagtggttccatattccaa 871
Db 624 CATCAAGGATATCTGACATCCACTCGTACTCCCAATGATGATCTACCTACTCANA 683
Qy 872 tacaggaagtaaaagcaagacatgaggaactgtctctagtagtgcagtggaagcagttcg 931
Db 684 TGGTGGTGTGAAATTTGGAGCTCTCGAAACCCCTGTGATGAAATGCTGCTAAAGCTACTGTGAA 743
Qy 932 tgcatttgcgaacaaactagtaaaataccaggtatatacacatggccatgctcagaacact 991
Db 744 AGACTTGGC---TCACCTGACGGGACCAAGTACACATATGGCCCGGAGCTACACAT 800
Qy 992 ataccatgctcctgaggtggtggaagcattggtatgatttggtggaatcaaatatcg 1051
Db 801 CATCTGCTGCTGGGGGCTTAAAGACTGGCTTATGACCAAGATGATATTCCTT 860
Qy 1052 tacat 1056
Db 861 CACCT 865

RESULT 14
US-08-860-882A-77
; Sequence 77, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy Disk
 COMPUTER: IBM compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/860,882A
 FILING DATE: JUNE 23, 1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: DONALD J. BIRD
 REGISTRATION NUMBER: 25,323
 REFERENCE/DOCKET NUMBER: 9901/238653
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 861-3027
 TELEFAX: (202) 832-0944
 TELEX: 6174627 CUSH
 INFORMATION FOR SEQ ID NO: 77:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1059 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-860-882A-77

```

: Sequence 112, Application US/09171945
: Patent No. 6277599
:
: GENERAL INFORMATION:
: APPLICANT: Emery, Stephen
: APPLICANT: Copley, Clive Graham
: APPLICANT: Edge, Michael Derek
:
: TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said
: TITLE OF INVENTION: Antibody, and Their Therapeutic Use in an Adept System
: FILE REFERENCE: Monoclonal Antibody to CEA
: CURRENT APPLICATION NUMBER: US/09/171,945
: CURRENT FILING DATE: 1998-10-29
:
: PRIOR APPLICATION NUMBER: GB9703103.3
: PRIOR FILING DATE: 1997-02-14
: PRIOR APPLICATION NUMBER: GB9609405.7
: PRIOR FILING DATE: 1996-05-04
: PRIOR APPLICATION NUMBER: PCT/GB97/01165
: PRIOR FILING DATE: 1997-04-29
:
: NUMBER OF SEQ ID NOS: 131
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 112
:
: LENGTH: 1870
:
: TYPE: DNA
:
: ORGANISM: Artificial Sequence
:
: FEATURE:
:
: OTHER INFORMATION: Description of Artificial Sequence: humanized
:
: US-09-171-945-112

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2002, 15:21:57 ; Search time 1764.15 Seconds

(without alignments)
12034.527 Million cell updates/sec

Title: US-09-980-881-1

Perfect score: 1573
Sequence: 1 agaaaattgctgttggaatg.....aaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

BSF:*
1: em_estba:*
2: em_estchum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	656	41.7	1446	11 AK004045	AK004045 Mus muscu
2	647	41.1	889	9 AL536069	AL536069 Mus muscu
3	622.8	39.6	780	10 BG616950	BG616950 602615625
4	615.6	39.1	806	10 BG567264	BG567264 602589745
5	604.2	38.4	788	10 BG618629	BG618629 602646064
6	581.2	36.9	747	10 BG566051	BG566051 602582652
7	579.2	36.8	725	10 BG618239	BG618239 602645967
8	561.8	35.7	738	10 BG616456	BG616456 602614618
9	555.2	35.3	735	9 AV646979	AV646979 AV646979
10	532.2	33.8	750	10 BG618813	BG618813 602646186
11	523.8	33.3	691	10 BG568240	BG568240 602587149
12	510.4	32.4	623	9 AV694058	AV694058 AV694058
13	509.4	32.4	583	9 AV693037	AV693037 AV693037
14	509.4	32.4	602	9 AV692032	AV692032 AV692032
15	507.4	32.3	621	9 AV698425	AV698425 AV698425
16	501.2	31.9	601	9 AV658390	AV658390 AV658390
17	491.6	31.3	643	9 AV688323	AV688323 AV688323

C	18	468	29.8	589	9	AV719484	AV719484
C	19	466	29.6	558	9	AV662298	AV662298
C	20	465.4	29.6	611	10	N63924	N63924
C	21	465	29.6	603	9	AM950700	AM950700
C	22	464	29.5	586	9	AV719308	AV719308
C	23	461.8	29.4	688	9	AV651709	AV651709
C	24	460.2	29.3	495	9	AI271641	AI271641
C	25	460.2	29.3	585	9	AV720894	AV720894
C	26	460.2	29.3	537	10	AA700377	AA700377
C	27	457.8	29.1	631	10	N98450	N98450
C	28	450.4	28.6	462	9	AV681868	AV681868
C	29	442	28.1	576	9	AV720000	AV720000
C	30	441.6	28.1	623	9	AV658178	AV658178
C	31	438.6	27.9	586	9	AV719997	AV719997
C	32	427.4	27.2	455	10	W88434	W88434
C	33	425.2	27.0	575	9	AL536068	AL536068
C	34	422.4	26.9	593	10	W89198	W89198
C	35	420.6	26.7	747	10	BG565100	BG565100
C	36	418.4	26.6	450	9	AI076434	AI076434
C	37	417.4	26.5	958	9	AI322676	AI322676
C	38	417	26.5	853	10	BI144644	BI144644
C	39	414.4	26.3	852	10	BG569281	BG569281
C	40	409.2	26.0	1022	10	BI332659	BI332659
C	41	405.6	25.8	544	9	BE235838	BE235838
C	42	397.8	25.3	530	9	AV720567	AV720567
C	43	396	25.2	580	9	AI182113	AI182113
C	44	396	25.2	589	9	AA968307	AA968307
C	45	394.4	25.1	661	9	AI529872	AI529872

ALIGNMENTS

RESULT	1	AK004045	1446 bp	mRNA	linear	HMC 19-JUN-2002
LOCUS	AK004045	Mus musculus 18 days embryo whole body cDNA, RIKEN full-length				
DEFINITION	AK004045	enriched library, clone:1110032P04:carboxypeptidase B2 (plasma), full insert sequence.				
ACCESSION	AK004045	AK004045.1	GI:12835067			
VERSION	AK004045					
KEYWORDS	AK004045	HTC, CAP trapper.				
SOURCE	AK004045	Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	AK004045	Mus musculus				
REFERENCE	AK004045	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.				
AUTHORS	AK004045	1 (sites)				
TITLE	AK004045	Carninci, P., and Hayashizaki, Y.				
JOURNAL	AK004045	High-efficiency full-length cDNA cloning				
MEDLINE	AK004045	Meth. Enzymol. 303, 19-44 (1999)				
PUBMED	AK004045	99279253				
REFERENCE	AK004045	10349636				
AUTHORS	AK004045	2 (sites)				
TITLE	AK004045	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.				
JOURNAL	AK004045	Normalization and subraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
MEDLINE	AK004045	Genome Res. 10 (10), 1617-1630 (2000)				
PUBMED	AK004045	2049374				
REFERENCE	AK004045	11042159				
AUTHORS	AK004045	3 (sites)				
TITLE	AK004045	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, Y., Yonekura, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuyama, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.				
JOURNAL	AK004045	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer				
TITLE	AK004045	Genome Res. 10 (11), 1757-1771 (2000)				

Db 985 TACTGTTTCCCTATTCCTATACAGAAAGCAAGACAGACAGAAAGTGTCTCTAG 1044
 QY 913 tagccagtgagagcagtcgtgctattgacaaactagtaaaataccaggatatacatg 972
 Db 1045 TGCCAGCAGCAGGACGATTCGTCGATTAAGATTATTAATAAAACACAGGATACACAGC 1104
 QY 973 gcatcgtcgaagaaccttactacgtctcgtgaggtgagggagatgatcatcattc 1032
 Db 1105 GCAGTGGCTCAGAAAGATTATATCTAGCTCTGAGCTTCGACGATTTGATATGATT 1164
 QY 1033 tggagcacaataatctcgtt----- 1051
 Db 1165 TGGGCAATCAAAATTTGCTTTACAAATGAGCTCCGAGATACAGCAGATACGATTTCTGC 1224
 QY 1052 -----tacatcaaacacacactgtagaagaagcttgcgcgtgctctctaaatag 1100
 Db 1225 TGCCCTGAGAGATATCAATCAACCCAGCTTTGTCAGAAAGCTTTGGCCGACATCTCTAAATAG 1284
 QY 1101 ctggagatgcatatgaatgattatgccccgattttatcattcgtctcgatttc 1160
 Db 1285 TTTGGCATGTCATCAGAAACACTTAATGCCC-----TAACTCTCCGCTCATTTATTTT 1338
 QY 1161 aattactgattccagcaagaacacatcatgcatcagat--tattttaattatcc 1218
 Db 1339 ATTATATGATTTTCAGCAACACTTAACCTTGTCATAGCTTTCAAGTTGAATCAGTTTCC 1398
 QY 1219 gtagtttgataaagat 1236
 Db 1399 TTGTTTGTGTTGAAGAT 1416

 RESULT 2
 AL536069 889 bp mRNA linear EST 13-FEB-2001
 LOCUS AL536069 LRI_F1013_Fbrn1 Homo sapiens cDNA clone CS0DF022YE21 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL536069
 VERSION AL536069.1 GI:12799562
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 889)
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1. 889
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF022YE21"
 /clone_lib="LRI_F1013_Fbrn1"
 /dev_stage="Pooled tissue from post conception fetuses (20
 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dt) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"
 255 a 186 c 192 g 255 t 1 others
 BASE COUNT
 ORIGIN

Query Match 41.1%; Score 647; DB 9; Length 889;
 Best Local Similarity 86.7%; Pred. No. 1,3e-85;
 Matches 771; Conservative 1; Mismatches 6; Indels 111; Gaps 1;

 QY 28 gcagccttgacgtctctgtgaccattgtctctctctctgagcaacagctctgcgttc 87
 Db 1 GCAGCCTTGACAGTCTCTGTACCCATTGTCTCTCTGTGACACAGATGCTCCGTTTC 60
 QY 88 agaagtgccaaagttctagctgctctctctctctctctctctctctctctctctctct 147
 Db 61 AGAGTGGCCAAAGTTCTAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
 QY 148 atcttactacaacatgagattgtctctgagcccggttaacagctgacattatgta 207
 Db 121 ATCTTACTACAACATATGAGATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
 QY 208 agaaaaaaccaagtcatttttttgaatgatcatgtagtgcgaatgtaaaagccatt 267
 Db 181 AGAAAAAACCAAGTCCATTTTGTAAATGATCTGATGTCGACAAATGTGAAGCCCAT 240
 QY 268 taaatgtagcaggaattccatgcatgctctctgagcaagctggaagatcttaaac 327
 Db 241 TAAATGAGAGGGAATTCACAGCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
 QY 328 agcagattcccaagacagtcagccccgagcctccgcatgctactatgaacagatc 387
 Db 301 AGCAGATTTCCCAAGCAGACAGTACGCCGCCGCTCCGACATCTGATGACAAAGTATC 360
 QY 388 actactcaaatgaaatctatctcttgatagaaattataactgagaagcactcgtatgc 447
 Db 361 ACTCACTAAATGAATATCTTCTTGATGATGATTAATACGAAAGCAGCATCTATATGC 420
 QY 448 ttacaaaatccacattgtagctcctcatcttgagaagtaaccactatggtttaagttt 507
 Db 421 TTACAAAATCCACATTGGATTCCTCATTTGAAAGTACCCACTTATGTTTAAAGTTT 480
 QY 508 ctggaagaagaacaaacagccaaatgcatatgtagtctgtagaattccatgccaag 567
 Db 481 CTGAAAAAGAACAGCAGCAGCAAAATGCAATGATGATGATGATGATGATGATGATGAT 540
 QY 568 aatgactctctccgct 608
 Db 541 AATGATCT 600
 QY 609 ----- 608
 Db 601 TAAATGGCAATATACCAATCTCTGAGGCTTGATTTCTATGTTATGCCGTGTTA 660
 QY 609 -----aatgaatgtagaagaacgcttct 636
 Db 661 ATGTGATGTTATGACTACTCATGAGAAAAAGATCGAATGAGAGAAAGACGTTCTT 720
 QY 637 tctatggaacaatcatctgcatctggaacagcccgatgaacttctctcaaacact 696
 Db 721 TCTATGGGAACAAATCATGATTCATTTGGAACAGCCGAAATAGCAATCTTCTCTCTCT 780
 QY 697 ggttgaggaagtgatccagcttcctcatgctgctggaacactactgagacttattctg 756
 Db 781 GGTGTGAGGAAGTGCATCCAGTTCTCTCATGCTCTGGAACCTTACTGTGACTTTATCTG 840
 QY 757 agtcaagaccagaagtgaaagcagtgtagttctcttgaagaagaatat 805
 Db 841 AGTCAGAACCCAGAGTGAAAGCAGCTGCTACTGTTCTTGAGAAAGAAAT 889

 RESULT 3
 BG616950 780 bp mRNA linear EST 18-APR-2001
 LOCUS BG616950 602615625F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4734247 5'
 DEFINITION mRNA sequence.
 ACCESSION BG616950
 VERSION BG616950.1 GI:13668321

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgs.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1597 row: b column: 08
High quality sequence stop: 754.

FEATURES
Source
Location/Qualifiers
1..780
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4734247"
/lab_host="NIH_MGC_76"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgccctggcc); Site_2: SfiI (ggcgccatagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 237 a 147 c 146 g 250 t
ORIGIN

Query Match 39.6%; Score 622.8; DB 10; Length 780;
Best Local Similarity 91.5%; Pred. No. 4,6e-82;
Matches 703; Conservative 0; Mismatches 12; Indels 53; Gaps 2;

QY 850 aataggtttccatctctctacacgaagtaaaagaacatgaggaactgtctc 909
|||||
Db 1 AATAGTGTTCATATTCCTATACAGAAATTAAGCAACATGAGAACTGTCTC 60
|||||

QY 910 tagtagccagtagaagaatctgctatgacaaaactagtaaaatccagatatacc 969
|||||
Db 61 TAGTAGCCAGTAGAAGACAGTTCGTGCTATGAGAAAATTAAGTAATAACAGGTATACAC 120
|||||

QY 970 atggccatggctcagaacactatatacctagctcccgaggtgaggaagatgcatat 1029
|||||
Db 121 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCGAGAGTGGGACATTTGATCTATG 180
|||||

QY 1030 atttgagcatcaaatatcgt----- 1050
|||||

Db 181 ATTGGGATCATCAAAATATTCGTTTACAAATTCGAGATACGGGACATACGATTCCT 240
|||||

QY 1051 -----ttacatcaaacaccctctgagagaagctttgacgctgctctaaa 1097
|||||

Db 241 TGCCTCCGAGAGCGTTACATCAAAACCTCTGAGGAAGCTTTGCGCGTCTCTTAAA 300
|||||

QY 1098 tagcttgacatgcataggaatgtttaatgcccctgattatcatctgcttcgtat 1157
|||||

Db 301 TAGCTTGACATGTCATAGGAATGTTTATGCCCCCTGATTTATCATCTGCTTCGAT 360
|||||

QY 1158 ttttaattacgattcccgacgaacaaatcatgtatcacgatatctttaagtttacc 1217
|||||

Db 361 TTTAATTACTGATTCAGCAAGACCAAAATCATTTGTATCATTTTAAAGTTTATTC 420
|||||

QY 1218 cgtagtttgataaaagatttccatctctctggtctcgcagagaacataaagtct 1277
|||||
Db 421 CGTAGTTTGATPAAAGATTTTCCATTCCTGTTCTCTGACAGAACCTAATTAATGCT 480
|||||

QY 1278 acttgcattaaagagactaggtgtcatctcttttccctttaaataaattgttaa 1337
|||||
Db 481 ACTTGCATTAAGGACAGACTAGGTTGATGCTTTTACCTTT-AAAAAAATTGTAA 539
|||||

QY 1338 aagctcagttacactcttctctctgatttcagcttgtagtaagcactaaagaact 1397
|||||
Db 540 AAGTCTAGTACTACTTTTCTTTTATTTTGACAGCTTTGACATCTCAAGCAAGT 599
|||||

QY 1398 ttgcagctttgactagcactcctcaagaagtttaatacaagaatcctcagctgcat 1457
|||||
Db 600 TTCGAGTTTGAAGTACCATCTCAAGCAAGTTTAAATCAATGATCATCTCAGCTGATCAT 659
|||||

QY 1458 tggatctctctcaacaaaggaaggtgtgtcgaagtaacataaagattctgtccaaa 1517
|||||
Db 660 TGGATCTTACTCAACAAAGGAGGAGGTGTGAGAGTACATTAAAGATTTCGTCCAAA 719
|||||

QY 1518 ttccaataattctctctctccttctctctctctctctctctctctctctctctct 1565
|||||

Db 720 TTTTCAATTAATTTCTGTTGCTTCCTTACTTAAAAAATTTTAAAAA 767
|||||

RESULT 4
BG567264 806 bp mRNA linear EST 10-APR-2001
LOCUS 602589745F1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4723690 5',
DEFINITION mRNA sequence.
ACCESSION BG567264 GI:13574917
VERSION BG567264.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgs.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1583 row: j column: 11
High quality sequence stop: 772.

FEATURES
Source
Location/Qualifiers
1..806
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4723690"
/lab_host="NIH_MGC_76"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgccctggcc); Site_2: SfiI (ggcgccatagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 239 a 158 c 162 g 247 t
ORIGIN

Query Match	39.1%	Score 615.6	DB 10	Length 806
Best Local Similarity	90.7%	Pred. No. 5.1e-81		
Matches 724	Conservative 0	Mismatches 14	Indels 60	Gaps 4
QY 614	aatgtgagaagaacacgcttcttctatgacgaacatcatlgtacatcggaacagacctga	673		
Db 1	AATGTGGAAGAAAGAACCTTCTTCTATGCGACAACTATTCGATCGGACAGACTTGAA	60		
QY 674	tageaaccttgtctccaaacacatggtgtgaaaggtgcatccagttctctatgtcga	733		
Db 61	TAGGAACCTTGTCTTCCAAACAACTGTGTGAGGAAGGTGCATCCAGTCTCTGCTCGGA	120		
QY 734	aacctactgtgtgactttatctctgagtcgacacgaagtgaagcagatggtgctagtctt	793		
Db 121	AACCTACTGTGTGACTTATTCCTGATCTGAAACCAAGATGGAAGGACGTGAGTGTCTT	180		
QY 794	gagaagaatatcaaccagatlaaagaatatacagacatgcatctatctatctccagcat	853		
Db 181	GAGAAGAATATCAACCAAGATTTAAAGCATACATCCATGCATCTTATATCTCCACATAT	240		
QY 854	agtgttccatattctctatacagaagaiaaagcaaaagccatgaggaactgtctagt	913		
Db 241	AGTGTTCATATTCCTATTCATACGAAGTAAAGCAAGAACCATGAGGAAGTCTCTAGT	300		
QY 914	agccagatgaagcagtgctgtctatttgacaacaaactgtataaaataccaggtatacaatg	973		
Db 301	AGCCAGTGAACAGTTCCTGTATTGACAAAACCTGTAAATATACAGATATACACATGG	360		
QY 974	ccatggtccagaacacttatccctatcctgctccctggaggtgggagcatttgatcattt	1033		
Db 361	CCATGGCTGAGAAACCTTATACCTATGCTCTGAGAGGTGGGACGATGTGATATGATTT	420		
QY 1034	gggacatcaaatatcgt-----	1050		
Db 421	GGGCATCAAAATATTCGTTTACAATGAACTTCGATACGAGATACGAGATTCCTGCT	480		
QY 1051	-----ttacatcaaacceccactgtagagaagcctttgcgcgtgtcctaaatagc	1101		
Db 481	GCCGGAGGCTTACATCAACCAACCCACTGTAGAGAAACCTTTGCCGCTGCTCTAAATATAGC	540		
QY 1102	ttggaatgcatatagaaatgtttaatgcccctgatttatacttgcctcgtatatta	1161		
Db 541	TTGGATGTCATATGGAATGTTAAATGCCCCGATTTATCATTTGCTCCGTATTTTAA	600		
QY 1162	atttactgattccagaacagacccaatcatctgcatcagattatltttaagtttaccgta	1221		
Db 601	ATTTACTGATTCACACAGACCAACCAATCATTTGATGAGATTTATTTTAAGTTTATCCGTA	660		
QY 1222	gttttgaataaagaatttctcatctcctgtgtctgtcagagaacctaataagtgctact	1280		
Db 661	GCTTTTGAATTAAGATTTTCCATATTCCTGTGTTGTGAGAGAACCTTAATAGTGA--	718		
QY 1281	ttgcatatgaagcgagactaagggttcatatgtcttcttacccttlaaaaaaaatltgaaag	1340		
Db 719	TTGGATTAAGGCAACATGAGGTTGATGCTTTTAAACCTTTAAAAAATGT---AAAG	773		
QY 1341	tctagtacctactttt 1358			
Db 774	TCTAGTACTAATTTCTT 791			
RESULT 5				
LOCUS	BG618629	788 bp	mRNA	linear
DEFINITION	602646006AFL NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',			EST 18-APR-2001
ACCESSION	BG618629			
VERSION	BG618629.1	GI:13670000		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 788)	NIH-MGC	http://mgc.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
Contact: Robert Strausberg, Ph.D.	Email: c9apbs-remail.nih.gov	Tissue Procurement: CLONETECH Laboratories, Inc.	cDNA Library Preparation: CLONETECH Laboratories, Inc.	DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: http://image.llnl.gov	Plate: LICM1630	row: 0	column: 15	High quality sequence stop: 760.
Location/Qualifiers	1..788	/organism="Homo sapiens"	/db_xref="taxon:9606"	/clone="IMAGE:4767974"
/clone.lib="NIH.MGC.76"	/lab_host="DH10B (TI phage-resistant)"	/note="Organ: Liver; Vector: pNCR-LIB (Clontech); Site: 1: SfiI (ggcgctccgcgcgc); Site: 2: SfiI (ggccatcttgcc); 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATGACC-3' and 3' adaptor sequence: 5'-ATCTGAGCGCGGAGGGCGGACATG-dT(30)-BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."		
BASE COUNT	222 a	157 c	169 g	240 t
ORIGIN				
Query Match	38.4%;	Score 604.2;	DB 10;	Length 788;
Best Local Similarity	99.5%;	Pred. No. 2.4e-79;		
Matches 606;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
1	agaaaattgctgttggagatgaagacgtttgagacgttcgagcttgcattgaccattgttctc	60		
49	acaaaattgctgttggagatgaagacgtttgagacgttcgagcttgcattgaccattgttctc	108		
61	tcctgtagcagatgcttcttcgcttccacagatgtagcgaagttcctaagcttctctctaga	120		
109	tcctgtagcagatgcttcttcgcttccacagatgtagcgaagttcctaagcttctctctaga	168		
121	cctctagcgaattcaagttctcaaggaatcttactacaacatagatgttctctctgac	180		
169	cctctagcgaattcaagttctcaaggaatcttactacaacatagatgttctctctgac	228		
181	agccgagtaacagcttaccctatttggaaagaaaacaaagttccatttttttggaaatgat	240		
229	agccgagtaacagcttaccctatttggaaagaaaacaaagttccatttttttggaaatgat	288		
241	ctgattgtagacaatgtagaaagcccatltaaatgtagcgaatctccatgtagtcttgc	300		
289	ctgattgtagacaatgtagaaagcccatltaaatgtagcgaatctccatgtagtcttgc	348		
301	tgagcagagctgtaaatcttattcaaacgagaatttccagagcagatgaccccgag	360		
349	tgagcagagctgtaaatcttattcaaacgagaatttccagagcagatgaccccgag	408		
361	ccctcgacatcgctactatgaacagatctacatcactaaatgaatcttatttggatagaat	420		
409	ccctcgacatcgctactatgaacagatctacatcactaaatgaatcttatttggatagaat	468		
421	tataactgagagagatcctctgatatgcttaaaaaaatccacgttggatcctcatttggaga	480		
469	ttatataactgagagagatcctctgatatgcttaaaaaaatccacgttggatcctcatttggaga	528		
481	agtaccacatcctatgtttttaaaggttttcggaaagaaacaaacagcccaaaatgtccatat	540		

Accession	Sequence	Position
Dd	AGTACCCTCTAAGTTTAAAGGTTTCTGAAAAAGAACAAAGCCAAAAATCCCAT	588
Qy	ggatttactgttgatccatgcccagagaaatgaaatcctcctgcttctgctgagttca	600
Dd	GGATTGACTGTGGAAATCCATCCCAAGAAATGGAACTCTCCTGCTTTCTGCTTGAGTTCA	648
Qy	taagccata 609	
Dd		
Dd	TAGGCCATA 657	

RESULT	6
BG566051	
LOCUS	747 bp mRNA linear EST 10-APR-2001
DEFINITION	60258265F01 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710564 5',
ACCESSION	BG566051 BG566051.1 GI:13573704
VERSION	EST.
KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	human

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (pages 1 to 747)	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	NIH-MGC http://mgc.nci.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)	Contact: Robert Strausberg, Ph.D.

FEATURES	source
Tissue Procurement: CLONETECH Laboratories, Inc.	
CDNA Library Preparation: CLONETECH Laboratories, Inc.	
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)	
DNA Sequencing by: Incyte Genomics, Inc.	
Clone distribution: wgc clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
Plate: LLCM1549	row: 9 column: 13
High quality sequence stop: 731.	
Location/Qualifiers	1..747

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4710564"
/clone_lib="NH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1; SfiI (ggcgccctcgagc); Site:2: SfiI (ggcgcatatggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCCATTTATGGC-3' and 3' adaptor sequence 5'-ATTCTACAGCCCGAGCGGCCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.8 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NH_MGC library."

```

Query Match	36.98;	Score 581.2;	DB 10;	Length 747;
Best Local Similarity	99.28;	Pred. No. 5.6e-76;		
Matches 605; Conservative	0;	Mismatches 3;	Indels 2;	Gaps 2;

1 agaaatctcgtctggatgagcttcgcagcctgcagtcctctgtacccatgttctct 60
 |||||
 4 AGAAATTCCTGTTGGGATGAGCGTTTGCAGCGCTTCCTGATCCCATTTG-TCCTC 62
 |||||

61 tctgtgagcagcatgtctcttcgcgtccacagatgacccaattctagctactcttcctaaaa 120

63 TCCTGACAGCATGCTTCGCGTTTCAGAGTCGCCAAGTCTTAGCTGCTTCTCTATAA 122

121 cctctaggcaagtccaagtctctacagaatcttactacaacatatgaattatctctgac 180

Db 123 CCTTAGGCAATTAAAGTTACACGAATCTTACTACAACTATGAACATGTTCTCGGC 182

QY 181 agccgtaacagctgaccta tggaaagaaaacaagtcacatttttgtaatgcat 240

Db 183 AGCCGTAACAGCTACCTTATGTGAAGAAAAACAAGTCATTTTGTGTAATGCAT 242

QY 241 ctgagtcgacaatctgagaaagcccatctaaatgtgagcggaaattccatgcatgtcttcgc 300
 |||||
 Db 243 CTGATGTCGACCAATGCGAAAGCCCATTTAAATGAGCGGAATTCATGCGATGCTCTTGC 302

QY 301 tggcagacgtggaagatcttattcaacagcagatttccaacgacacagtcccccag 360
|||||
Db 303 TGGCAGACGTGGAGATCTTATTCACAGCAGATTTCACAGACAGTACGCCCCGAG 362

OY 361 cctccgcatcgtaactgaacagtatcaactcactaaatgaatatctattcttgatagaat 420
|||||
Db 363 ccmcccacmccmccmcc

26 CCCTCCGATCGTACATGAGCAGTATCACTATAAGAAATCTATCTTGATATGAAT 422
 Oy 421 ttataactgagagcatcctgatgatgttacaataatccacatrgatcctcattgaga 480
 |||||

Db 423 TTATAACTGAGAGGCATTCCTGATATGCTTACAAAAATCCACATTTGGATCTCTCAATTTGAGA 482

QY 481 agtaccactctatgttt-taaagttctctggaagacaaacagccaanaatgcata 539

Db	483	AGTACCAC	CTATGTTT	CTAAAGTTT	CTG	AAAAAG	AACAAG	CAGCC	AAAAAT	GC	CATA	542
QY	540	tggattgact	gtgnaatcc	atgcgcga	agaagaat	tgatctctc	ctcttctc	cttctg	cttctg	cttctg	cttctg	599

Db	Seq	Start	End	Score	Value
543	TGGATTGACTGTGGAATCCATGCCAGAGAAATGGATCTCTCCCTTTCTGCTTGTGGTTC	1	602	100	100
600	atagagccata	609			

Db 603 ATAGCCCATTA 612

RESULT	7	
BG618239		
LOCUS		
BG618239		
725 bp		
MPNA		
1400000		
ECM 18-1000-2001		

DEFINITION	6026645967F1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4767376 5', 3'-UTR
ACCESSION	U00000.1
VERSION	B618239
KEYWORDS	EST.
SOURCE	human.

ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 725)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: L1C6M1629 row: f column: 17
High quality sequence stop: 723.

```

FEATURES
source
Location/Qualifiers
1. 725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4/67376"
/clone_lib="NH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccatcattggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor

```

sequence: 5'-CACGGCCATTATGCCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGACGGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 203 a 153 c 149 g 220 t
ORIGIN

Query Match 36.8%; Score 579.2; DB 10; Length 725;
Best Local Similarity 99.2%; Pred. No. 1.1e-75;
Matches 603; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 agaaatgtctgttgatgaagcttgcagccttcagctcctgtacccattgtctc 60
DB 48 AGAAATGTCTGTGGATGAAGCTTGCAAGCTTGCAAGCTTGCAAGCTTGCT 107
QY 61 tctgtgagcagcalttctcgcgttcagagtgagcgaatttagctgtcttcctga 120
DB 108 TCTGTGAGCAGCATGCTCTGCGCTTTCAGAGTGCCAAAGTTCTAGCTCTCTGAGAA 167
QY 121 cctctgagcagatltcaagttctacagatcttatacaatagagattgtctctg 180
DB 168 CCGTAGGCAAGTTCACAGATCTTACATACATATAGAGA-TGTTCTCTGGC 226
QY 181 agccggaagcagctgagcctattgtgagagaaagacagctcatttttctgaatgat 240
DB 227 AGCCGGAACAGCTGACCTTA-TGTGAAGAAAAACAAGTCATTTTGTGAATGAT 285
QY 241 ctgattgcagcaatlygaaagccattaaatlygagcgaattcattgcagtgcttc 300
DB 286 CTGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCGGAAATTCATGCAAGTCTTGC 345
QY 301 tggcagacgctggaagatcttattcaacagagatttccaaagacagtaagcccccag 360
DB 346 TGGAGATGTGGAAGATCTTATTCACAGAGATTTCCAAAGACAGTCAGCCCCGAG 405
QY 361 cctcgcagatgtaactagacagatcactcaactaaatgaatcattctctgtgagaa 420
DB 406 CTTCCGATGTGATGAGACAGATCACTACATGAATGAATCTATTTGGATAGAT 465
QY 421 ttatactgagagcagcctgatatgcttcaaaaatccacattgagctcctcattaga 480
DB 466 TTATTAATGTGAGAGCATCTGATGCTTCAAAAATCCACATGATGATCCATTTAGA 525
QY 481 agtaaccacatctatgttttaaaagttcttgaagaaagaaagccaaaatgcatat 540
DB 526 AGTACCCACTGATGTTTAAAGTTCTGGAAGAAAGCAAGCAAAAATGCCATAT 585
QY 541 ggaatgagctgtgagcagcagcagatgagatcctcctgtcttctgtctgtgttca 600
DB 586 GGAATGACTGTGGAATCCATGCCAGAAATGAGATCTCTCCGCTTCTGCTTGGTTCA 645
QY 601 taagccat 608
DB 646 TAGGCCAT 653
RESULT 8
LOCUS BG616456 738 bp mRNA linear EST 18-APR-2001
DEFINITION 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
ACCESSION BG616456
VERSION BG616456.1 GI:13667827
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutheria; Chordata; Cranialia; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 738)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@pds-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LNCM1595 row: C column: 05
High quality sequence stop: 666.

FEATURES
source
1. 738
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4733500"
/lab_host="NIH-MGC-76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1;
SfiI (ggcgccctggcg); Site: 2: SfiI (ggcgattagcgc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGCCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGACGGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 209 a 161 c 153 g 215 t
ORIGIN

Query Match 35.7%; Score 561.8; DB 10; Length 738;
Best Local Similarity 99.6%; Pred. No. 3.9e-73;
Matches 563; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 45 gtaccatgttctccttctgtgagcagatgtcttcgagcagatgagcgaatttcta 104
DB 1 GTACCATGTGCTCTCTGTGAGCAGATGTCTTCGGTTTCAGAGGCCAAGTTCTA 60
QY 105 gctgccttctcagaaacctttagcgaattcaagtlctacagaatcttactacaacat 164
DB 61 GCTGCTCTCTCTAGAACCTCTAGCAGAGTCAAGTTCTACAGAACTTACTAACAATAT 120
QY 165 gagattgtctctgagcagcgttaacagctgaccttattgtgaagaaacaagtcacat 224
DB 121 GAGATTGTCTCTGCGACCGGTAAACGCTTATTTGGAAGAAAAACAAGTCCAT 180
QY 225 ttcttgaatgacatcgtatgctgacaaatgtgaagccatttaaatgtgagcgaatt 284
DB 181 TTTTGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 285 ccacgagctgtctgtcgtgagcagcgtggaagatcttatacaacagcagatttccaagc 344
DB 241 CCATGCAAGTGTCTGCGACGAGCTGCAATGATGATGATGATGATGATGATGATGAT 300
QY 345 acaatcagccccgagcctcgcagctcgtatcattgaacagatcactcaatgaatc 404
DB 301 ACAATGACCCCGAGCGCTCGCATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 405 tattctgagatgataataacagagagcagcctgtatgcttatacaaaatccacatt 464
DB 361 TATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 465 ggaatcctcatttgaagatccacactcattgtttaaagtttctgtgaaagaaaca 524
DB 421 GGATCTCTATTTGAAGAGTACCCACTCATGTTTAAAGGTTTCGGAAGAAAGCAAGCA 480
QY 525 gccaaaatgcatatgattgactgtggaatcattgcagagatgattgtctccgct 584
|||||

DB 481 GGCNAAAATGCAATATGATGACTGTGTGATTCATGCGAGAAATGATCTCTCTGCT 540

QY 585 ttctgctgtgtgttcataaggccata 609
|||||

DB 541 TTCTGCTTGTGGTTCATAGGCCATA 565

RESULT 9

AV646979 735 bp mRNA linear EST 15-JAN-2002

LOCUS AV646979 GLC Homo sapiens cDNA clone GICATD06 3', mRNA sequence.

DEFINITION AV646979

ACCESSION AV646979

VERSION AV646979.1 GI:9867993

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 735)

AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,D., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z., and Han,Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

COMMENT Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919 (ex.45) Fax: 86-21-50801922 Email: hanzg@hgc.sh.cn

FEATURES

source This clone is available at CHGC in Shanghai.

Location/Qualifiers

1..735

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="GICATD06"

/clone_lib="GLC"

/issue_type="corresponding non cancerous liver tissue"

/dev_stage="Adult"

/lab_host="SOLR"

/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 200 a 158 c 172 g 205 t

ORIGIN

Query Match 35.3%; Score 555.2; DB 9; Length 735;
Best Local Similarity 98.4%; Pred. No. 3.6e-72;
Matches 571; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 31 gecttcagtccttgcacatgtctctctgtgtgagcagatgtctgcgttcacaga 90
|||||

DB 5 GCGTTCGCGCTGTGACCATTTGTTCTCTGTGAGAGAGAGATGCTTCGGCTTCACA 64
|||||

QY 91 gtggcgaagttcagctgtctcttcacgaaccttaagcgaagttcaagttcacagaatc 150
|||||

DB 65 GTGGCCAACTTCTAGCTGCTCTTCTCTAGAACCTTAGGCACTTCAAGTTCTACAGAAATC 124
|||||

QY 151 ttactacaacatatagatgtctctctgcgcagcggtaacacagctacacctttgtgaaga 210
|||||

DB 125 TTACTACAAACATATGAGATGTCTCTGCGACCGCGTAAACAGTACCTTATGTGAAGA 184
|||||

QY 211 aaaaacagtcacatttttttgaatgcatcgtatgtgcgaatgtgaagccattaa 270
|||||

DB 185 AAAAAACAAGTCATTTTGTGAATGCAATGATGATGCAACAATGTGAAGCCCAATTAA 244
|||||

QY 271 atgtgagcgaatcctcagcagtgctctgtgcgcagcagtggaagatttatcaacgc 330
|||||

DB 245 ATGTGAGCGGAATTCATGCAGTGTCTTCTGCGACAGCTGGAAGATCTTATTCACAGC 304

QY 331 agatttccaacgacacagtcagcccccagccctccgcacatctgatacagtcatact 390
|||||

DB 305 AGATTTCACAGACACAGTACAGCCCGACGCTCCCGATCGGACATGAAACATATACCT 364
|||||

QY 391 cactaaatgaatctatctcttgatagaattataactgaaggaacctctgatatagctta 450
|||||

DB 365 CACTTAATGAAATCTATTTCTTGATGAAATTTTAACCTGAGAGCATCTCGATATCTTAA 424
|||||

QY 451 caaaatccacattgatactcctcatttgaagaagaccacactatgttttaa- gtttct 509
|||||

DB 425 CAAAAATCCAAATTCATTCCTCATTTGAGAGATCCCATCTATGTTTAAAGGGTCT 484
|||||

QY 510 ggaagaagaacaaacagccaaatagccatatgattgactgtggaatccatgcagaaga 569
|||||

DB 485 GGAAGAAGAACAAACACGCCAAATAATGCCATATGATGACTGTGGAATCATGCAGAGAA 544
|||||

QY 570 tggatctctctgtcttctgtctgtgtgtgttcataaggccata 609
|||||

DB 545 TGGATCT 584
|||||

RESULT 10

BG618813 750 bp mRNA linear EST 18-APR-2001

LOCUS BG618813

DEFINITION 602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:476785 5', mRNA sequence.

ACCESSION BG618813

VERSION BG618813.1 GI:13670184

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 750)

AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgsrbs@mail.nih.gov

Tissue Procurement: CLONETECH Laboratories, Inc. cDNA Library Preparation: CLONETECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LINC630 row: k column: 22 High quality sequence stop: 622.

FEATURES

source Location/Qualifiers

1..750

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:476785"

/clone_lib="NIH_MGC_76"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: liver; Vector: pDNR-LTR (Clontech); Site_1: SfiI (ggcgctggcc); Site_2: SfiI (ggccattagcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCATTAATGGC-3' and 3' adaptor sequence: 5'-ATTCTAAGGCGGAGCGCGGACATG-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 219 a 146 c 171 g 213 t

ORIGIN

Query Match 33.8%; Score 532.2; DB 10; Length 750;
Best Local Similarity 98.0%; Pred. No. 8.1e-69;

Matches 581; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

QY 9 gctgttggaagaagcttgcagccttcgagcttcgttacccattgtctctctgtgag 68
 DB 2 gctgttggaagaagcttgcagccttcgagccttcgttacccattgtctctctgtgag 61
 QY 69 caacatgtcttcgagcttcgagccttcgagccttcgttacccattgtctctctgtgag 128
 DB 62 caacatgtcttcgagcttcgagccttcgagccttcgttacccattgtctctctgtgag 121
 QY 129 caagttcaagttctacagaatcttactacacatagatgtctctctgcagccgta 188
 DB 122 caagttcaagttctacagaatcttactacacatagatgtctctctgcagccgta 181
 QY 189 acagctgcaccttattgtgaagaanaaacaagtcacattttttgtgaatgcacgtatgc 248
 DB 182 acagctgcaccttattgtgaagaanaaacaagtcacattttttgtgaatgcacgtatgc 241
 QY 249 gacacatgtgaagaagccatttaaatgtgagcgaattccatgcagctgtctgtgagac 308
 DB 242 gacacatgtgaagaagccatttaaatgtgagcgaattccatgcagctgtctgtgagac 301
 QY 309 gtggaagaattcttcaagaagcagatttcaagaagcagcagtcagcccgagccctgcga 368
 DB 302 gtggaagaattcttcaagaagcagatttcaagaagcagcagtcagcccgagccctgcga 361
 QY 369 tctgactactgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 428
 DB 362 tctgactactgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 421
 QY 429 gagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 486
 DB 422 gagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 481
 QY 487 cactc-tatgttttaaaaggttctt-ggaagaagaacaagcagcagcagcagcagcagc 544
 DB 482 cactc-tatgttttaaaaggttctt-ggaagaagaacaagcagcagcagcagcagcagc 541
 QY 545 tgaactgtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 597
 DB 542 tgaactgtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 594

RESULT 11
 LOCUS BG568240 691 bp mRNA linear EST 10-Apr-2001
 DEFINITION BG568240.1 GI:13575893
 mRNA sequence.
 ACCESSION BG568240
 VERSION BG568240.1 GI:13575893
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 691)
 AUTHORS NIH-MGC
 TITLE NIH-MGC
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.lnl.gov
 Plate: LICM1563 row: a column: 14
 High quality sequence stop: 691.
 Location/Qualifiers
 1..691
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone="IMAGE:4715797"
 /clone_1id="NIH_MGC_76"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:
 SfilI (ggccgcccggcc); Site_2: SfilI (ggccattatggc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CAGGCGCATTTAGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGCGCGCGCATATG-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC Library."

BASE COUNT 204 a 143 c 144 g 200 t
 ORIGIN

Query Match 33.3%; Score 523.8; DB 10; Length 691;
 Best Local Similarity 98.5%; Pred. No. 1.4e-67;
 Matches 539; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 3 aaattgctgtgttggaagaagcagcagcagcagcagcagcagcagcagcagcagcagc 62
 DB 4 AAAATGCTGTGGAGATGAAGCTTTGACGCTTGCAGCTTGTGATCCCATTTGCTTTC 63
 QY 63 tgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 122
 DB 64 TGTGAGCAGCATGTCTTCCGCTTTCAGAGTGTGCAAGTCTAGCTGCTCTTCTGTAAGC 123
 QY 123 tctaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 182
 DB 124 TCTGAGCAGCATGTGCAAGCTTTGACGCTTGTGATCCCATTTGCTTTC 183
 QY 183 ccggttaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 242
 DB 184 CCGGTAAACAGCTGACCTTA-TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
 QY 243 gatctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 302
 DB 243 GATGTCGACATGTGGAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTGCTG 302
 QY 303 gcagacgtlgaagaatcttactcaagaagcagcagcagcagcagcagcagcagcagc 362
 DB 303 GCAGACGTGGAAGATCTTATTCACAGCAGATTTCCAAAGCAGCAGCAGCAGCAGC 362
 QY 363 tccgcatcgtctacttgaacagcagcagcagcagcagcagcagcagcagcagcagcag 422
 DB 363 TCCGATCTGATCTATGAACAGTATCTACTCACTAAATGAAATCTATTTCTGTGATGA 422
 QY 423 ataactgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 482
 DB 423 ATAACTGAGAGCATTCCTGATATCTTACAAAATCCCATTTGATGATCTTGTGAGAG 482
 QY 483 taacctactatgtltaaaaggttcttgcgaaagaacaagcagcagcagcagcagcag 542
 DB 483 TACCACATCTATGTTTAAAGTTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 542
 QY 543 attagc 549
 DB 543 ACTCAAT 549

RESULT 12
 LOCUS AV694058 623 bp mRNA linear EST 16-Jan-2002
 DEFINITION AV694058 GKC Homo sapiens cDNA clone GKCADB09 5', mRNA sequence.
 ACCESSION AV694058
 VERSION AV694058.1 GI:10295921
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```
OY 3 aaattgctgttggaatgaagcttgcagccttcgacgtctctgttaccattgtctcttc 62
   |||||||
Db 1 AAAATTGCTGTGGATGTAAGCTTTCAGCTTCGAGCTCTGTACCATTTGTTCTCTTC 60
   |||||||
OY 63 tctgagcagcatgtctctgcgttcagagtgagcgaagttcagctgtctctcctagaacc 122
   |||||||
Db 61 TGTGAGCAGCATGTCTTCGCTTCAGAGTGCCCAAGTTCTAGCTGCTCTCTCTAGAACCC 120
   |||||||
OY 123 tctaggcaagttcgaagttctacagaaactctactacacaatagagatgttctctgcag 182
   |||||||
Db 121 TCTAGGCAAGTTCAAGTTCTACAGAACTTACTACACATATAGAGATTGTTCTTGCCAG 180
   |||||||
OY 183 ccggttaacagctgcaccttatgtgaagaataaaacaaagtcacatttttgttaaatgcatc 242
   |||||||
Db 181 CCGGTAAACAGCTGACCTTATGTGAAAGAAAACAAAGTCCATTTTGTGTAATGCACTCT 240
   |||||||
OY 243 gatgtcgaaatgtgaaagcccaattaaatgtgagcggaaatccatgcagtgtctgtcg 302
   |||||||
Db 241 GATGTCGACAAATGTGAAAGCCCAATTAATGTGAGCGGAATTCATGCAAGTGTCTTGCTG 300
   |||||||
OY 303 gcaagcgtggaagatcttaattcaacagcagatttccaacgacagtcagcccccgagcc 362
   |||||||
Db 301 GCAGACGTGGAAGATCTTATTCACACGCAATTTCCACACACAGCTCAGCCCCGAGCC 360
   |||||||
OY 363 tccgacatcgtactatgaacagatcacctcacctaaatgaatctatctctgtgataaatt 422
   |||||||
Db 361 TCCGATCTGACTATGAAACAGTATCAGCTCACTAAATGAAATCTATCTGTGATAGAAATT 420
   |||||||
OY 423 ataactgagagcaccctcgtatagtcttaaaaaaatccacattgcatcctcattgagaag 482
   |||||||
Db 421 ATAACTGAGAGGCAATCCGATATGCTTACAAAAATCCACATTGGATCCCATTTGAGAAG 480
   |||||||
OY 483 taaccactctatgtctttaaaggttctcgtgaagaagaa 521
   |||||||
Db 481 TACCCACTCTATGNTNTAAAGSTTCTGGAACCAAGCCA 519
   |||||||
```

Search completed: September 18, 2002, 16:32:00
Job time: 4203 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:43:08 ; Search time 23.83 Seconds

(without alignments)
584.937 Million cell updates/sec

Title: US-09-980-881-2

Perfect score: 1911

Sequence: 1 MKICSLAVLPVLFCEQHV.....IKYFPTSNPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	692	36.2	415 1 CBPB_RAT	P19223 ratius norv
2	641	33.5	417 1 CBPB_HUMAN	P15086 homo sapien
3	629.5	32.9	401 1 CBPB_PIG	P09955 sus scrofa
4	628.5	32.9	417 1 CBPB_MOUSE	P15089 mus musculu
5	627	32.8	416 1 CBPB_CANFA	P55261 canis famli
6	619.5	32.4	417 1 CBPB_HUMAN	P15088 homo sapien
7	580	30.4	306 1 CBPB_BOVIN	P00732 bos taurus
8	553	28.9	417 1 CBP2_RAT	P19232 ratius norv
9	546	28.6	309 1 CBPB_RAT	P21961 ratius norv
10	533.5	27.9	419 1 CBPB_BOVIN	P00730 bos taurus
11	529.5	27.7	419 1 CBPB_RAT	P00731 ratius norv
12	508	26.6	417 1 CBPB_HUMAN	P48052 homo sapien
13	489.5	25.6	419 1 CBPB_HUMAN	P48052 homo sapien
14	483	25.3	421 1 CBPB_HUMAN	P09142 homo sapien
15	404.5	21.2	304 1 CBPB_STIMI	P42788 stimulum vi
16	386	20.2	433 1 CBPB_ANGOA	O02350 anopheles g
17	369.5	19.3	303 1 CBPB_ASTFL	P04059 astacus flu
18	299.5	15.7	430 1 YHTT_YEAST	P29068 saccharomyc
19	277.5	14.5	424 1 CBPT_THEVU	P42091 anguilla an
20	271	14.2	451 1 CBPS_STRGR	P18143 streptomyce
21	251	13.1	434 1 CBPS_STRCP	P39041 streptomyce
22	118.5	6.2	707 1 ORCI_SCHPO	P54789 schizosacch
23	105	5.5	1255 1 PER2_HUMAN	O15055 homo sapien
24	100	5.2	376 1 YOGT_BACSU	P54497 bacillus su
25	99.5	5.2	445 1 TR5H_CHICK	P70080 gallus gall
26	98.5	5.2	448 1 TY3H_ANGAN	O42091 anguilla an
27	97	5.1	663 1 TAZI_SCHPO	P09005 schizosacch
28	95	5.0	444 1 TR5H_HUMAN	P17752 homo sapien
29	93.5	4.9	3587 1 SRFL_BACSU	P27206 bacillus su
30	91.5	4.8	465 1 STHA_ECOLI	P27306 escherichia
31	91	4.8	297 1 Y103_HUMAN	O15006 homo sapien
32	91	4.8	444 1 TR5H_RABIT	P17290 cycloclagus
33	91	4.8	982 1 ENV_SFV3L	P27399 simian foam

34	90	4.7	1451	1	SPT6_YEAST	P23615 saccharomyc
35	89.5	4.7	682	1	CDBA_BACUH	O86170 bacillus th
36	89	4.7	525	1	PUR9_BUCAI	P57143 b bifunctio
37	88.5	4.6	319	1	UN30_CAEL	P52906 caenorhabdi
38	88	4.6	454	1	CBPH_LOPAM	P37892 lophus ame
39	88	4.6	1150	1	C9EA_BACTA	O92n19 bacillus th
40	87	4.6	516	1	YKE7_YEAST	P36090 saccharomyc
41	86	4.5	523	1	STK_PYRHO	O57963 pyrococcus
42	85.5	4.5	396	1	ENP1_BACSH	O03415 bacillus sp
43	85.5	4.5	476	1	CBPH_RAT	P15087 ratius norv
44	85	4.4	481	1	TR5H_XENLA	O92142 xenopus lae
45	85	4.4	1047	1	RIRL_CHLTR	O84834 chlamydia t

ALIGNMENTS

```

RESULT 1
CBPB_RAT ID CBPB_RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-89034324; PubMed-3182872;
RA Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-Lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
EMBL; M23959; AAA40872.1; -
EMBL; M23947; AAA40872.1; JOINED.
EMBL; M23950; AAA40872.1; JOINED.
EMBL; M23952; AAA40872.1; JOINED.
EMBL; M23953; AAA40872.1; JOINED.
EMBL; M23954; AAA40872.1; JOINED.
PIR; A32129; A32129.
HSSP; P09955; INSA.
MEROPS; M14.003; -
InterPro; IPR003146; Propep_M14.
InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CROXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 13
FT PROPEP 14 108
FT CHAIN 109 415
FT METAL 174 174
FT METAL 177 177
FT METAL ZINC (BY SIMILARITY).
FT METAL ZINC (BY SIMILARITY).

```

FT METAL 302 302 ZINC (BY SIMILARITY).
 FT ACT_SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 171 184 BY SIMILARITY.
 FT DISULFID 243 266 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 SO SEQUENCE 415 AA; 47515 MM; 8BA0CCADE30B672 CRC64;

Query Match 36.2%; Score 692; DB 1; Length 415;
 Best Local Similarity 39.2%; Pred. No. 6,9e-51;
 Matches 150; Conservative 61; Mismatches 126; Indels 46; Gaps 7;

QY 1 MKLGSIAVPIVIFECQHFVAFSGOVLALPRTSRQVYVQLNTTYEIVLMQPVADL 60
 DB 1 MLLALVVALAHASEH---FDGNRYRVSVHGEDHVNLIQELANTKEIDFWKPDASAT 57
 QY 61 LTVKKOVHFEVNASDVNVKAKHLNAGICPSVYLADVEDLIQOQISNDYSPRASASY 120
 DB 58 QVRLTVDFVFKAKEDVADVDFEENDEYHREYVLSVNRALLESQFDSHT---RASGHSY 114
 QY 121 EYHSLNEIYSWIEETTERHPDMLTKIHGSSFEKYPYLVLKSGKDOTAKNAIWDGCI 180
 DB 115 TKYKMETIEMWIOQVATNDPDLVTOGVTGTEGRNMYVLKI-GKTRPKPAIFIDCGF 173
 QY 181 HAREWISPAFCIMFI-----GH-----NRMRK 203
 DB 174 HAREWISPAFCIMFVRAVTVYNOELHMKOLDELDFYLPVYNIDYVYTWTKDRMRK 233
 QY 204 NRSFYANNHICIGDNLNSFNYSKHCERGASSSCSEYCGLYPSEPEVKAASFLEARNI 263
 DB 234 TRSTMAGSSCLGVAPRNPNF-NAGMCEVAGASRSPCEYCGPAPASEKETALADFTINNL 292
 QY 264 NQKAYISMSYSQHTVFPYSTRSKSKDHEELSLVASEVRAIDKTSKFTRYTHGSE 323
 DB 293 STTKAVLTHSYSQMMLYPYSYKLPENYELNALVKGAKEL-ATLHGTKYTYGPGAT 351
 QY 324 TLVLAPGGDDMIYLDGIRKSYFT 346
 DB 352 TIYPAAGSDMSYDQIKISFT 374

RESULT 2
 CBBP_HUMAN STANDARD; PRT: 417 AA.
 AC P15086; O60834;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein) (PASP).
 GN CPB1 OR CPB OR PCPB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI-TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
 RC TISSUE-Pancreas;
 RX MEDLINE=92129345; PubMed=1370825;
 RA Yamamoto K.K., Foussette A., Chow P., Wilson H., el Shami S., French C.K.;
 RT "Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B.";
 RL J. Biol. Chem. 267:2575-2581(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Pancreas;
 RX MEDLINE=98182241; PubMed=9524066;
 RA Aloy P., Cataus L., Villegas V., Reverter D., Vendrell J., Aviles F.X.;
 RT "Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B.";

RL BIOL. Chem. 379:149-155(1998).
 RN [3]
 RP SEQUENCE OF 16-43.
 RC TISSUE-Pancreas;
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT "Purification and properties of five different forms of human procarboxypeptidases";
 RL Eur. J. Biochem. 179:609-616(1989).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
 CC peptide + L-Lysine(or L-arginine).
 CC -1- TISSUE SPECIFICITY: PANCREAS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL: M61057; AAA6973.1; -;
 DR EMBL: AJ224866; CA12163.1; -;
 DR PIR: S02812; S02812.
 DR PIR: A42332; A42332.
 DR HSSP: P09955; 1NSA.
 DR MEROPS: M14.003; -;
 DR MIM: 114852; -;
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR PRINTS: PR00765; CROBOXYPASA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 110
 FT CHAIN 111 417
 FT METAL 176 176
 FT METAL 179 179
 FT METAL 304 304
 FT ACT_SITE 378 378
 FT DISULFID 173 186
 FT DISULFID 245 268
 FT DISULFID 259 273
 FT CONFLICT 16 16
 FT CONFLICT 17 17
 FT CONFLICT 37 37
 FT CONFLICT 208 208
 FT CONFLICT 245 245
 SO SEQUENCE 417 AA; 47366 MM; BBIQF212D830305E CRC64;

Query Match 33.5%; Score 641; DB 1; Length 417;
 Best Local Similarity 35.3%; Pred. No. 1.4e-46;
 Matches 135; Conservative 71; Mismatches 132; Indels 44; Gaps 7;

QY 3 LGSIAVPIVIFECQHFVAFSGOVLALPRTSRQVYVQLNTTYEIVLMQPVADL 61
 DB 1 MLLALVVALAHASEH---FDGNRYRVSVHGEDHVNLIQELANTKEIDFWKPDASAT 57
 QY 61 LTVKKOVHFEVNASDVNVKAKHLNAGICPSVYLADVEDLIQOQISNDYSPRASASY 120
 DB 58 QVRLTVDFVFKAKEDVADVDFEENDEYHREYVLSVNRALLESQFDSHT---RASGHSY 114
 QY 121 EYHSLNEIYSWIEETTERHPDMLTKIHGSSFEKYPYLVLKSGKDOTAKNAIWDGCI 180
 DB 115 TKYKMETIEMWIOQVATNDPDLVTOGVTGTEGRNMYVLKI-GKTRPKPAIFIDCGF 173
 QY 181 HAREWISPAFCIMFI-----GH-----NRMRK 203
 DB 174 HAREWISPAFCIMFVRAVTVYNOELHMKOLDELDFYLPVYNIDYVYTWTKDRMRK 233
 QY 204 NRSFYANNHICIGDNLNSFNYSKHCERGASSSCSEYCGLYPSEPEVKAASFLEARNI 263
 DB 234 TRSTMAGSSCLGVAPRNPNF-NAGMCEVAGASRSPCEYCGPAPASEKETALADFTINNL 292
 QY 264 NQKAYISMSYSQHTVFPYSTRSKSKDHEELSLVASEVRAIDKTSKFTRYTHGSE 323
 DB 293 STTKAVLTHSYSQMMLYPYSYKLPENYELNALVKGAKEL-ATLHGTKYTYGPGAT 351
 QY 324 TLVLAPGGDDMIYLDGIRKSYFT 346
 DB 352 TIYPAAGSDMSYDQIKISFT 374

ID	CPBA_PIG	STANDARD	PRT	401 AA.
Db	177	AREMISPAFCQMFREAVRRTYGRGIQVTELLNKLPFLVPLVNLIDGITYTWTSSRRFRKT	236	
Qy	205	RSFYANNHCICTDITDINSFVSKHMCCEGASSSCSEYTGLYPSEPEPKAVASFLLRRNIN	264	
Db	237	RSTHTGSSCICITDNPNNRF-DAGMCEIGASRNPCDEFTYCGPAASEKFTKLADPFRMILS	295	
Qy	265	QIKAVISMHSIOMHIVPPYSTTRKSKDHEELSLVASEAVRALDTSKNTRYTHGHSCT	324	
Db	296	SIRAKVILTHSYDOMYIVSYAVYKIGENNALNALAKATVKEL-ASLHGKRYGYGPQATY	354	
Qy	325	LYLAPGGDDWIVDYGIRKYSFT	346	
Db	355	IYPAAGSGSDMAYDQIGIRYSFT	376	
RESULT	3			
CPBA_PIG		STANDARD	PRT	401 AA.
AC	P09955			
DT	01-MAR-1989	(Rel. 10, Created)		
DT	15-JUL-1998	(Rel. 36, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Carboxypeptidase B precursor	(EC 3.4.17.2).		
GN	CPB.			
OC	Sus. scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_Taxid=9823;			
RM	[1]			
RP	SEQUENCE OF 1-103.			
RX	MEDLINE=91208150; PubMed=2018774;			
RA	Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,			
RA	Aviles F.X.;			
RT	"Analysis of the activation process of porcine procarboxypeptidase B			
RT	and determination of the sequence of its activation segment.";			
RL	Biochemistry 30:4082-4089(1991).			
RN	[2]			
RP	PRELIMINARY SEQUENCE OF 1-38.			
RX	MEDLINE=85279427; PubMed=4026847;			
RA	Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;			
RT	"Sequential homologies between procarboxypeptidases A and B from			
RT	porcine pancreas.";			
RL	Biochem. Biophys. Res. Commun. 130:97-103(1985).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).			
RX	MEDLINE=91114690; PubMed=1989878;			
RA	Coll M., Guasch A., Aviles F.X., Huber R.;			
RT	"Three-dimensional structure of porcine procarboxypeptidase B: a			
RT	structural basis of its inactivity.";			
RL	EMBO J. 10:1-9(1991).			
RN	[4]			
RP	STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.			
RX	MEDLINE=91027767; PubMed=2223783;			
RA	Vendrell J., Wider G., Aviles F.X., Wuehlich K.;			
RT	"Sequence-specific 1H NMR assignments and determination of the			
RT	secondary structure for the activation domain isolated from			
RT	pancreatic procarboxypeptidase B.";			
RL	Biochemistry 29:7515-7522(1990).			
RN	[5]			
RP	STRUCTURE BY NMR OF ACTIVATION PEPTIDE.			
RX	MEDLINE=91114693; PubMed=1989879;			
RA	Vendrell J., Billeter M., Wider G., Aviles F.X., Wuehlich K.;			
RT	"The NMR structure of the activation domain isolated from porcine			
RT	procarboxypeptidase B.";			
RL	EMBO J. 10:11-15(1991).			
RN	[6]			
RP	STRUCTURE BY NMR OF ACTIVATION PEPTIDE.			
RX	MEDLINE=93044373; PubMed=1422143;			
RA	Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,			
RA	Huber R., Wuehlich K.;			
RT	"Comparison of the NMR solution structure with the x-ray crystal			
RT	structure of the activation domain from procarboxypeptidase B.";			

BL	J.	Biomol.	NMR	2:1-10(1992).
CC	-I-	CATALYTIC ACTIVITY:	Peptidyl-L-lysine(or L-arginine) + H ₂ O =	
CC	-I-	peptide + L-lysine(or L-arginine).		
CC	-I-	SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE		
CC	-I-	ZINC CARBOXYPEPTIDASE FAMILY.		
CC	-I-	DATABASE: NAME=Worthington enzyme manual;		
CC		WWW="http://www.worthington-biochem.com/manual/C/COB.html".		
DR	PDB:	B29181; B29181.		
DR	PDB:	1PBA; 31-OCT-93.		
DR	PDB:	INSA; 24-DEC-97.		
DR	MEROPS:	M14.003; -.		
DR	InterPro:	IPR003146; Proped_M14.		
DR	InterPro:	IPR000834; Zn_carbopept.		
DR	pfam:	PF02246; Proped_M14; 1.		
DR	pfam:	PF00246; Zn_carbopept; 1.		
DR	PRINTS:	PR00765; CRBOXYPTASEA.		
DR	PROSITE:	PS00132; CARBOXYPEPT_ZN_1; 1.		
DR	PROSITE:	PS00133; CARBOXYPEPT_ZN_2; 1.		
KW	Hydrolase;	Carboxypeptidase; Metalloprotease; zinc; zymogen;		
KW	3d-structure.			
FT	PROPEP	1	95	ACTIVATION PEPTIDE.
FT	CHAIN	96	401	CARBOXYPEPTIDASE B.
FT	DISULFID	158	171	
FT	DISULFID	230	253	
FT	DISULFID	244	258	
FT	METAL	161	161	ZINC.
FT	METAL	164	164	ZINC.
FT	METAL	289	289	ZINC.
FT	ACT SITE	341	341	
FT	ACT SITE	363	363	NUCLEOPHILE.
FT	STRAND	14	17	
FT	HELIX	20	31	
FT	TURN	32	32	
FT	HELIX	43	45	
FT	STRAND	50	52	
FT	HELIX	61	69	
FT	TURN	70	72	
FT	STRAND	75	76	
QO	SEQUENCE	401 AA;	45713 MW;	53129AFI59A26348 CRC64;

[illegible]

```

RESULT 4
CBPC_MOUSE
ID CBPC_MOUSE STANDARD: PRT: 417 AA.
AC P15089;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Serafin W.E.;
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT novel member of the carboxypeptidase gene family."
RL J. Biol. Chem. 264:20094-20099(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05118; AAA37369.1; -
DR PIR: A34487; A34487.
DR HSSP: P09955; 1NSA.
DR MEROPS: M14.010; -.
DR MGD: MGI:88479; Cp33.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carbopept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carbopept; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
SQ SEQUENCE 417 AA; 48790 MW; AAB300A068D1BA6D CRC64;

Query Match 33.9%; Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%; Pred. No. 1.5e-45;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;
OY 6 LAVLVPIVLFCEQHFVAFQSGOVLALPRTSRQVQLQNTLTTEIVLMQPTADLIYKK 65
DB 7 MAVIYTTLAIPVH---FDREKVFVKLQNEKHSVYKLNLTQSLDEWYDAIHDAIVN 63
OY 66 KQVHEFVNASVDNKKALNNSGIPCSVTLADVEDLIQOQIS-NDYSPRASASTYEQYH 124
DB 64 MTPDPRVSEKESQTIQSTLEQHKIHYELLHDLQELIEKQFDVDELAGRHS---YAKYN 120

```

```

OY 125 SLNEIYSWIEFIERHPDMLTKIHIGSFPEKYLPLYLAKYSGEQAKNAIWDGCIHARE 184
DB 121 DMDKITSWTEKMLEKHPDEMYSIRIKISTVEDNPLVYKTI-GRKDGERRAIFMDCGHARE 179
OY 185 WISPAFLMEI-----GHN-----RWMRKNSF 207
DB 180 WISPARCFQWFGATKSYGKNKIMTKLIDRMNFYLPVENVDGCIYMSWTQDMKRNRSR 239
OY 208 YANNHICIGDLDNENFYSKNCWCEGASSSCSEFYGLYSEPEYKAVASFLRRINQIK 267
DB 240 NQNSTCIGIDLNNF-DVSWDSSPNTNKPCLNVRGPAPESEKRAVNFIRSHLSIK 298
OY 268 AVISMHSYQHIYFPYSYTRSKSKDHEELSLVASEAVRIDKTSNRTYTHGSGSETLYL 327
DB 299 AVTFHSYSQMLIPYGYFKLPENHQLAKVARIATDAL-STRETRIYSPASTIYK 357
OY 328 APGGGDWYIDGIRKSF 345
DB 358 TSGSSLDWYDGLIKHTF 375

RESULT 5
CBPB_CANFA
ID CBPB_CANFA STANDARD: PRT: 416 AA.
AC P55261;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE membrane associated protein) (ZAP47).
GN CPBL.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_Taxid=9615;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Fukuoka S.-I.;
RL Submitted (Nov-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2)O -
CC peptide + L-lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D78348; BAAL1366.1; -
DR HSSP: P09955; 1PBA.
DR MEROPS: M14.003; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carbopept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carbopept; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 110 416 CARBOXYPEPTIDASE B.
FT METAL 175 175 ZINC (BY SIMILARITY).
FT METAL 178 178 ZINC (BY SIMILARITY).
FT METAL 303 303 ZINC (BY SIMILARITY).
FT ACT_SITE 377 377 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 244 267 BY SIMILARITY.

```



```

Db 177 ARHWISAPFQWVYQATKTYGKRNKTKLDRNPFYILFVNVNDGYTSMWTKNRMRKN
QY 205 RSEFYANNHCIGTDLNNSNFVSKHMCCEGASSSCSEYTCGLYPESEPEVKAVASFLRRN 264
Db 237 RSKNNSKICIGTDLNRRNF-NASWNSIPNTNDPCADNRGSAPESEKTKAVTFINSHLN 295
QY 265 QIKAYISMHSYSHIYFPYSTRSKSDHEELVASEAVRAIDKSKNTRYHGHGSET 324
Db 296 ELIVYTFHYSQMLFPYGYTSKLPNHBHLAKVAKIGTDVL-STRYETRYIYGIEST 354
QY 325 LYLAPGCGDDWIDYDGIKYSF 345
Db 355 IYPISSSSLDMAVDGIKHTF 375

```

```

RESULT 7
CBPB_BOVIN STANDARD; PRT: 306 AA.
AC P00732:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B (EC 3.4.17.2).
GN CPB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN [2]
RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RA MEDLINE=74260705; PubMed=4833744;
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RA Schmid M.F., Herrloft J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RL J. Mol. Biol. 103:175-190(1976).
RN [4]
RP ACTIVE SITE.
RA MEDLINE=70007159; PubMed=5344132;
RT "Isolation and sequence of peptides at the active center of bovine
carboxypeptidase B.";
RL J. Biol. Chem. 244:5246-5253(1969).
RN [5]
RP ACTIVE SITE.
RA MEDLINE=73061487; PubMed=4565668;
RT "Identification of a glutamic acid at the active center of bovine
carboxypeptidase B.";
RL J. Biol. Chem. 247:7864-7869(1972).
RN [6]
RP CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
peptide + L-Lysine(or L-arginine).
RN [7]
RP SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
ZINC CARBOXYPEPTIDASE FAMILY.
CC PIR: A00912; CPBOB.
DR MEROPS: M14_003; -.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF00246; Zn_carboxypept; 1.

```

```

DR PRINTS: P00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT DISULFID 149 163
FT METAL 149 163
FT METAL 66 66
FT METAL 69 69
FT METAL 194 194
FT ACT_SITE 246 246
FT ACT_SITE 268 268
FT ACT_SITE 346 346
SQ SEQUENCE 306 AA; 34612 MW; C329D2655C4A172 CRC64;

```

```

Query Match 30.4%; Score 580; DB 1; Length 306;
Best Local Similarity 43.6%; Pred. No. 1,2e-41;
Matches 115; Conservative 44; Mismatches 65; Indels 40; Gaps 5;

```

```

QY 120 YEQYSLNMEIYSEWIEFTRHPDMLTKHIGSSFERYPLVYKSGKEQTAKNALWIDCG 179
Db 6 YEKYNWETIEAMTEQVASENDLISRAIGTFILGNTIYLKV-GKQSKNPAYFMDCG 64
QY 180 IAREWISAPFCLMPT-----GH-----NRMR 202
Db 65 FAREWISAPFQWVYQATKTYGRIHMTFELDKLDFYLPVNVNDGYITWTNRMR 124
QY 203 KNRSEFYANNHCIGTDLNNSNFVSKHMCCEGASSSCSEYTCGLYPESEPEVKAVASFLRRN 262
Db 125 KTRSTRAGSSCTGTDLNRNF-DAGMCSIGASNNPCSEYTCGSAASEKESKAVADFLRNH 183
QY 263 INQIKAYISMHSYSHIYFPYSTRSKSDHEELVASEAVRAIDKSKNTRYHGHGS 322
Db 184 ISSIRAYVLIHYSQMLFPYGYTSKLPNNVELMTLAKAVKKL-ASLHGTTISYGPGA 242
QY 323 EFLYAPGCGDDWIDYDGIKYSF 346
Db 243 TTIYPAAGSDMAVDGIKHTF 266

```

```

RESULT 8
CBP2_RAT STANDARD; PRT: 417 AA.
AC P19222:
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Pancreas;
RA MEDLINE=89034323; PubMed=3182871;
RA Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.B.,
RA Graf M., Rutter W.J.;
RT "A novel rat carboxypeptidase, CPA2: characterization, molecular
cloning, and evolutionary implications on substrate specificity in
the carboxypeptidase gene family.";
RL J. Biol. Chem. 263:117828-117836(1988).
RN [2]
RP SEQUENCE OF 131-143 FROM N.A.
RA MEDLINE=95386501; PubMed=7657630;
RA Normant E., Gros C., Schwartz J.C.;
RT "Carboxypeptidase A isoforms produced by distinct genes or
alternative splicing in brain and other extrapancreatic tissues.";
RL J. Biol. Chem. 270:20543-20549(1995).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA MEDLINE=92105124; PubMed=1761558;

```


QY 263 INQIATVSMHSYSHIVPYSTRSKSKDEELSLVASEAVRAIDKTSKRTYTHGHGS 322
 Db 186 LNSIKAVYTFHSYSQMLFPYGYTIKLEPNQDILKVARIRTDVL-SSRYETRIYIGPIA 244
 QY 323 ETLVAPGSGDDWYDGIKTSF 345
 Db 245 STIKTSGSSLDWAVDGIKHTF 267

RESULT 10
 CBPA BOVIN
 ID CBPA BOVIN STANDARD; PRT: 419 AA.
 AC P00730:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase A precursor (EC 3.4.17.1).
 GN CPA.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91151335; PubMed=1998496;
 RT "Cloning and nucleotide sequence of a bovine pancreatic
 RT preprocarboxypeptidase A cDNA."
 RL Biochem. Biophys. Res. Commun. 175:110-116(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Pancreas;
 RC MEDLINE=96096549; PubMed=8522204;
 RA Geo J.H., Kim K.H., Choi K.Y.;
 RT "Cloning, sequencing and expression of the gene encoding a major
 RT allelic preprocarboxypeptidase A from bovine pancreas."
 RL Gene 165:333-334(1995).
 RN [3]
 RP SEQUENCE OF 111-417.
 RA MEDLINE=71155183; PubMed=5102489;
 RA Bradshaw R.A., Walsh K.A., Neurath H.;
 RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
 RT chymotryptic peptides of the cyanogen bromide fragment F-I."
 RL Biochemistry 10:938-950(1971).
 RN [4]
 RP REVISIONS TO 138 AND 141.
 RA MEDLINE=72138789; PubMed=5143102;
 RA Petra P.H., Hermanson M.A., Walsh K.A., Neurath H.;
 RT "Characterization of bovine carboxypeptidase A (Allan)."
 RL Biochemistry 10:4023-4025(1971).
 RN [5]
 RP SEQUENCE OF 17-120.
 RA MEDLINE=89150306; PubMed=3147705;
 RA Wade R.D., Haas G.M., Kumar S., Walsh K.A., Neurath H.;
 RT "The amino acid sequence of the activation peptide of bovine pro-
 RT carboxypeptidase A."
 RL Biochimie 70:1137-1142(1988).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
 RA MEDLINE=83294519; PubMed=6887246;
 RA Reed D.C., Lewis M., Lipscomb W.N.;
 RT "Refined crystal structure of carboxypeptidase A at 1.54-A
 RT resolution."
 RL J. Mol. Biol. 168:367-387(1983).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
 RA MEDLINE=96003618; PubMed=7556081;
 RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
 RT "The three-dimensional structure of the native ternary complex of
 RT bovine pancreatic procarboxypeptidase A with proproteinase E and
 RT chymotrypsinogen C.";

RL EMBL J. 14:4387-4394(1995).
 RN [8]
 RP VARIANT ALLELIC.
 RA MEDLINE=69283620; PubMed=5817619;
 RA Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
 RT "Identification of the amino acid replacements characterizing the
 RT allelic forms of bovine carboxypeptidase A."
 RL Biochemistry 8:2762-2768(1969).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
 CC amino acid.
 CC -1- SUBUNIT: MONOMER. THE ZMOGEN IS SECRETED AS A TERNARY COMPLEX
 CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
 CC PROPROTEINASE E.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -1- DATABASE: NAME=Worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M61851; AAA30426.1; -;
 DR EMBL: M61851; AAA30427.1; -;
 DR EMBL: Z33906; CAA83955.1; -;
 DR PIR: JN0126; CPBOA.
 DR PIR: A31406; A31406.
 DR PDB: 3CPA; 15-JAN-87.
 DR PDB: 4CPA; 22-OCT-84.
 DR PDB: 5CPA; 15-JAN-87.
 DR PDB: 6CPA; 15-OCT-91.
 DR PDB: 7CPA; 31-JAN-94.
 DR PDB: 8CPA; 31-JAN-94.
 DR PDB: 1CPB; 31-JAN-94.
 DR PDB: 1CPB; 15-OCT-94.
 DR PDB: 2CTB; 31-JAN-94.
 DR PDB: 2CTB; 31-JAN-94.
 DR PDB: 1ARL; 01-AUG-96.
 DR PDB: 1ARM; 17-AUG-96.
 DR PDB: 1BAV; 01-APR-97.
 DR PDB: 1YME; 12-FEB-97.
 DR PDB: 1CPX; 05-AUG-98.
 DR PDB: 1PYT; 27-JAN-97.
 DR MEROPS: M14.001; -;
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxypept; 1.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
 KW 3D-structure; Polymorphism.
 FT SIGNAL: 1
 FT PROPEP: 17
 FT CHAIN: 111
 FT METAL: 179
 FT METAL: 182
 FT METAL: 306
 FT ACT_SITE: 358
 FT ACT_SITE: 358
 FT ACT_SITE: 380
 FT DISULFID: 248
 FT VARIANT: 289
 FT VARIANT: 338
 FT VARIANT: 338
 FT CONFLICT: 415
 FT CONFLICT: 95
 FT CONFLICT: 199
 FT CONFLICT: 203
 FT CONFLICT: 224
 FT CONFLICT: 224

ACTIVATION PEPTIDE.
 CARBOXYPEPTIDASE A.
 ZINC.
 ZINC.
 ZINC.
 PROTON DONOR.
 NUCLEOPHILE.
 I -> V (IN ALLELIC VARIANT).
 E -> A (IN ALLELIC VARIANT).
 L -> V (IN ALLELIC VARIANT).
 S -> L (IN REF. 5).
 D -> N (IN REF. 3).
 D -> N (IN REF. 3).
 D -> N (IN REF. 3).
 D -> N (IN REF. 3).

FT	CONFLICT	232	232	0 -> E (IN REF. 3).
FT	CONFLICT	295	295	D -> N (IN REF. 3).
FT	TURN	114	116	
FT	TURN	119	120	
FT	HELIIX	125	138	
FT	TURN	140	142	
FT	STRAND	143	150	
FT	TURN	152	153	
FT	STRAND	156	162	
FT	STRAND	171	176	
FT	TURN	180	181	
FT	HELIIX	183	199	
FT	TURN	200	202	
FT	HELIIX	204	212	
FT	STRAND	214	218	
FT	HELIIX	223	231	
FT	TURN	232	232	
FT	TURN	234	235	
FT	STRAND	242	242	
FT	TURN	244	245	
FT	STRAND	249	249	
FT	TURN	253	254	
FT	STRAND	261	262	
FT	TURN	267	267	
FT	TURN	270	271	
FT	TURN	273	274	
FT	STRAND	275	275	
FT	TURN	280	281	
FT	HELIIX	284	296	
FT	STRAND	299	307	
FT	STRAND	311	314	
FT	TURN	324	325	
FT	HELIIX	326	344	
FT	STRAND	349	352	
FT	HELIIX	353	356	
FT	TURN	357	357	
FT	HELIIX	364	370	
FT	TURN	371	372	
FT	STRAND	375	381	
FT	TURN	388	389	
FT	HELIIX	393	416	
SO	SEQUENCE	419 AA;	47082 MW;	21B86407B3BFC452 CRC64;

[illegible]

```

FT METAL 306 306 ZINC (BY SIMILARITY).
FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 248 271 BY SIMILARITY.
FT CONFLICT 196 196 I -> V (IN REF. 1).
FT CONFLICT 261 263 FGM -> LGR (IN REF. 1).
FT CONFLICT 347 347 K -> E (IN REF. 1).
SQ SEQUENCE 419 AA; 47197 MW; BB002DICB9B7491 CRC64;

Query Match 27.7%; Score 529.5; DB 1; Length 419;
Best Local Similarity 34.5%; Pred. No. 3.4e-37;
Matches 134; Conservative 63; Mismatches 136; Indels 55; Gaps 10;

QY 2 KLSGLAVPIVLFCEQHFVAFSGOVLALPRTSRQOVQLNLTYYEIVL--WQPVTA 59
DB 3 RLILSLLEAVCGNEN---FVGHQVLRISSADEAQVKKELEDENLQDFWRDAR 58
QY 60 DLIVKKQVHEFNASDVNVAHLNVSIGPCVLLADVEDL---QQQISNDTVSPRAS 116
DB 59 AGIPIDVRFV---PSIQSVKAFLEYHGISTYIMIEDVQLLDEERKQMSAFQARALST 114
QY 117 ASY-YEQVHSLNFIYSWIEFETERHPDLRIKIHGSSFEKYPVLYLVKSGEQATKANM 175
DB 115 DSENYATYHTLDEIYERMDLLVAEHQVLVKIQIGMTFEGRPILHVKFS-TGGINPRATM 173
QY 176 IDGCIHAREKISPAFCIMFIGH-----N 198
DB 174 IDGCIHAREKISPAFCIMFIGH-----N 198
QY 199 RMRKRNSFEYANHCIGTDLNSNFVSKHMCESGSSCSEYCGLYPSEPEVKAVAF 258
DB 234 RMRKRNRHSGSLCYGVDPNRNW-DAGCGMACASSNPCEITRGFPNSEVEVKSIYDP 292
QY 259 LRRNIQIAVYSMHSYSGHIYEPYSTRSKSDHEELSLVASEAVRAIDKTSKNTRYTH 318
DB 293 VYSHGN-IRAFISIHYSQLLIYPGYTSEPPADQAELOLQAKSAVTAL-TSLHGKFKY 350
QY 319 GHGSETLYLAPGGCDWITDLGIKYSFT 346
DB 351 GSTIDITYQASGSTIDWYSQGIKYSFT 378

RESULT 12
CPB2_HUMAN
ID CPB2_HUMAN STANDARD; PRT: 417 AA.
AC P48052;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
RC TISSUE=Pancreas;
RX MEDLINE=95204457; PubMed=7896805;
RA Catusus L., Vendrell J., Aviles F.X., Carreira S., Pulgarer A.,
RA Billeter M.;
RT "The sequence and conformation of human pancreatic
RT procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT three-dimensional model.";
RL J. Biol. Chem. 270:6651-6657(1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98111000; PubMed=9450539;
RA Reverter D., Garcia-Saez I., Catusus L., Vendrell J., Coll M.,
RA Aviles F.X., Garcia-Saez I., Catusus L., Vendrell J., Coll M.,
RT "Characterisation and preliminary X-ray diffraction analysis of human
RT pancreatic procarboxypeptidase A2.";

```

```

RL FEBS Lett. 420:7-10(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98046021; PubMed=9384570;
RA Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT "The three-dimensional structure of human procarboxypeptidase A2,
RT deciphering the basis of the inhibition, activation and intrinsic
RT activity of the zymogen.";
RL EMBO J. 16:6906-6913(1997).
CC -I- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC residues.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U19977; AAA74425.1; -
DR PDB: IAYE; 13-JAN-99.
DR MEROPS: M14.002; -.
DR MIM: 600688; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CARBOXYPEPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW 3D-structure.
FT SIGNAL 1 16
FT PROPEP 17 112 POTENTIAL.
FT CHAIN 113 417 ACTIVATION PEPTIDE.
FT METAL 177 177 CARBOXYPEPTIDASE A2.
FT METAL 180 180 ZINC.
FT METAL 304 304 ZINC.
FT ACT_SITE 378 378 ZINC.
FT DISULFID 246 269 NUCLEOPHILE.
FT DISULFID 318 352
SQ SEQUENCE 417 AA; 46840 MW; 6D8107330A2DBC58 CRC64;

Query Match 26.6%; Score 508; DB 1; Length 417;
Best Local Similarity 32.5%; Pred. No. 2.2e-35;
Matches 128; Conservative 63; Mismatches 137; Indels 66; Gaps 11;

QY 19 HVE---AFOSGOVLALPRTSRQOVQLNLTYYEIVL--WQPVTA DLIVKKQVHEFN 73
DB 13 HIYCEITVGVQVLETVSNBQIKNLQDLAEOEHLQDFWKSFT---TPGETAHVRVP 68
QY 74 ASDVDNVAHLNVSIGPCVLLADVEDLQQQISNDTVSPRAS---YEQVHSLNFI 130
DB 69 FVNVQAVKVFLESQGIAYSIMIEDVQLLDKNEMLFNRRRSGNPNFCAVITLEIS 128
QY 131 SWIEETIRRHDMLEKIHGSSFEKYPVLYLVKVS-GEQAKNAIMWIDCGIHAREWISPA 189
DB 129 QEMDLVAEHGLVSKVNIYSSFPENPMVNLKFTSGD---KPAIMWDAGIHAEVWYQA 185
QY 190 FCLWFIIG-----HNRMWRKRNSFEYAN 212
DB 186 TALMTANKIVSDYGDKPSITSLDALDIFLLPYNPGIYFSGTKNMRKRTRSKVSGSL 245
QY 213 CIGTDLNSNFVSKHMCESGSSCSEYCGLYPSEPEVKAVASFLRRNIQIAVYSM 272
DB 246 CVGVDPNRNW-DAGFGGPGASSNPSCSDSYHGPSANSEVEVKSIYDFIKSH-GVKNAFIIL 303
QY 273 HSYSGHIYEPYSTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGG 332

```

```

Db 304 HSSQLMPEYKCTKRLDDFDLSEVQAQAASL-RSLHGTRKYKPGICSVIYQASGGS 362
QY 333 DMWYDLGIRKYSFTSN-----PVEKLKP 356
Db 363 IDMSYDGIKYSFAPFLRDTGRYGFLLPAPQILP 396

RESULT 13
CBP1_HUMAN
ID CBP1_HUMAN STANDARD: PRT; 419 AA.
AC P15085;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN CPA1 OR CPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA MEDLINE=93038569; PubMed=1417781;
RA Catasus L., Villegas V., Pascual R., Aviles F.X.,
RA Wicker-Planquart C., Pulgover A.;
RT "CDNA cloning and sequence analysis of human pancreatic
RT procarboxypeptidase A1."
RL Biochem. J. 287:299-303(1992).
RN [2]
RP SEQUENCE OF 17-42.
RX MEDLINE=89153096; PubMed=2920728;
RX Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT "Purification and properties of five different forms of human
RT procarboxypeptidases."
RL Eur. J. Biochem. 179:609-616(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X67318; CAA47732.1; -
DR PIR: S02810; S02810.
DR PIR: S29127; S29127.
DR HSSP: P00730; 1PRT.
DR MEROPS: M14.001; -.
DR MIM: 114850; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CARBOXYPEPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KM Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 110
FT CHAIN 111 419
FT METAL 179 179
FT METAL 182 182
FT METAL 306 306
FT ACT_SITE 358 358
FT ACT_SITE 380 380
ACT_SITE NOCLEOPHILE (BY SIMILARITY).

```

```

FT DISUPID 248 271 BY SIMILARITY.
SQ SEQUENCE 419 AA; 47140 MW; 439FAFFFAEE958B1 CRC64;

Query Match 25.6%; Score 489.5; DB 1; Length 419;
Best Local Similarity 32.7%; Pred. No. 8e-34;
Matches 131; Conservative 60; Mismatches 127; Indels 83; Gaps 13;

QY 3 LCSLAVLPVILFCEQHVFAFGSGVLAALPRTRQOVYLONLTYEIVL--WQ----- 55
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4 LVLVSLVLAAGVEKED----FVGHQVLRISVADQAQYKKELEDLEHLQDLDFMRGPAHP 59
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 --PYTADLVKKQKQVHFVNASDVDAKNAHNSGICSVLADVEDLI---QQQI----- 106
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 GSPIDV-----RVPF-----PSIQAVKIFLESHGISYETEMIEDVQSLDEEQEOMFAFR 108
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 107 ----SNDTVPASASVSEQYHSLNETYSMEETTERHPMLTKIHGSSFEKPLVYK 162
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109 SRARSTDFN-----YATHTLEELYDFDLVAENPHLVSKIQIGNYEGRPVYVK 161
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 163 VSGKEQTAKNAWIDCGIHAREWISPAFLMTI----- 195
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 162 FS-TGSKRRPAIMWDIGHSREWYTOASGVFAKKITQDYGDQDAFTALDITDIFLEIV 220
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 196 -----GHNRMRKNSFYANNHCIGTDLNSNFVSKHWCESGSSSCSEFTYGLY 245
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 221 TNPDGFAFTSTNMRMKRTSRHTAGSLCIGVDPRRN--DAGFGLSGASSNPCESTYHGKF 279
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 246 PESPEEVKAVAFPRRNINIKAVISMHSYQHVPEFYSTRSKRHELSUVAFAVR 305
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 280 ANSEVEKSVDFVKDGN--IKAFISHSYSQLMYPRGYKTEPVPQDDELQSKAAVT 338
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 306 AIDKTSKTRTYTHGSGSETLYLAFGGDDWYIDGIRKYSFT 346
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 339 AL-ASLGTGKFNKGIKAIYQASGSTITDWTYSGIKYSFT 378
| :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
CBP4_HUMAN
ID CBP4_HUMAN STANDARD: PRT; 421 AA.
AC G9U142;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN CPA4 OR CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99310537; PubMed=10383164;
RX Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT deacetylase inhibitors during differentiation of prostate epithelial
RT cancer cells."
RL Cancer Res. 59:2981-2988(1999).
CC -1- FUNCTION: Could be involved in the histone hyperacetylation
CC pathway.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF095719; AAF23230.1; -
DR HSSP: P48052; IAYE.

```

DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxypept; 1.
 DR PRINTS: PR00765; CARBOXYPEPTASE.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 DR HydroLase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
 FT SIGNAL 1 16
 FT PROPEP 17 113 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 114 421 CARBOXYPEPTIDASE A4.
 FT METAL 181 181 ZINC (BY SIMILARITY).
 FT METAL 184 184 ZINC (BY SIMILARITY).
 FT METAL 308 308 ZINC (BY SIMILARITY).
 FT ACT_SITE 382 382 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 250 273 BY SIMILARITY.
 SQ SEQUENCE 421 AA; 47379 MW; 91883D25B705BC6A CRC64;

Query Match 25.3%; Score 483; DB 1; Length 421;
 Best Local Similarity 32.0%; Pred. No. 2.8e-33;
 Matches 119; Conservative 63; Mismatches 130; Indels 60; Gaps 10;

QY 23 FQSGVLAALPRTSROVOVLQNLTTTEIVL--WQPVTA-----DLIVKKQVHEFVNAS 75
 DB 21 FFGDVLRLVNRNGDEISRLSQLVNSNMKLNPMKSPSSFNRPVVLVPSVSLQAF---- 76
 QY 76 DVDNKAHLNVSGIPCSVLLADVEDLI--QQISNDTVSPRASASY-YEOXSLNEIYS 131
 DB 77 -----KSLFRSGCLEAVAYTIEDLQALLNDEDMQHNQEGSSNNFNCAHSLAITH 131
 QY 132 WIEFTERHPDMLTKIHIGSSFEKYPPLVLYKVGKEQTAKNAIWDGCIHANEWISPAFC 191
 DB 132 EMDNIAADPDRLARRVKIKHSFENRPMYLYKEFTGKGVRRPAVWLNAGIHSSEWISQAFA 191
 QY 132 LM-----FI-----GHNRMRNRSEFYANNHCT 214
 DB 192 IWTARKIVSDVORDPATISILEKMDIFLIPVANDGYVTYTONRLMRTKRSNPGSSCT 251
 QY 215 GTDLNSNFVSKHMCESGSSSCSEYTCGLYPESEPEVAVASFLRRNINQIKAYISMHS 274
 DB 252 GADPRNRW-NASFAKGASDNCSEYHNGPHANSEVEVASYVDIQKHGN-FKGFIDLS 309
 QY 275 VSOHIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKTRTHGSGETLYIARPGSGD 334
 DB 310 YSQLMTYPGYSVKAPDAEELDKVARKALAKALASVS-GTEYVGPECTCTYVPASGSSID 368
 QY 335 WIYDGIKYSFT 346
 DB 369 WAYDNGIKRFAFT 380

RESULT 15
 CBP2_SIMV1 STANDARD; PRT; 304 AA.
 ID CBP2_SIMV1
 AC P42788;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
 OS Simulium vittatum (Black fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 OC Chironomidae; Simuliidae; Simulium.
 OC NCBI_Taxid=7192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Gut;
 RX MEDLINE=94093864; PubMed=8269093;
 RX Ramos A., Mahowald A., Jacobs-Jorena M.,
 RT "Gut-specific genes from the black fly Simulium vittatum encoding
 RT trypsin-like and carboxypeptidase-like proteins.";
 RL Insect Mol. Biol. 1:149-163(1993).

CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
 CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: L08481; AA018531.1; -.

DR HSSP: P48052; IAVE.

DR InterPro: IPR000834; Zn_carboxypept.

DR Pfam: PF00246; Zn_carboxypept; 1.

DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.

DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.

KW HydroLase; Carboxypeptidase; Metalloprotease; zinc.

FT NON_TER 1 1

FT METAL 58 58 ZINC (BY SIMILARITY).

FT METAL 61 61 ZINC (BY SIMILARITY).

FT METAL 184 184 ZINC (BY SIMILARITY).

FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).

FT DISULFID 125 148 BY SIMILARITY.

FT SEQUENCE 304 AA; 34849 MW; 2E6E3FF8A6A9144 CRC64;

Query Match 21.2%; Score 404.5; DB 1; Length 304;
 Best Local Similarity 33.0%; Pred. No. 7.4e-27;
 Matches 93; Conservative 52; Mismatches 86; Indels 51; Gaps 9;

QY 122 QXSLNEIYSWIEFTERHPDMLTKIHIGSSFEKYPPLVLYKVGKEQTAKNAIWDGCIH 181
 DB 1 QXHTLEIYSWIDRLVQEHREHPEVPGSKSYEGNEIRGVKYSYK--GNPVWVESNH 58
 QY 182 AREWISPA---FCL-----WFI-----GH-----NRMRKNRS 206
 DB 59 AREWISPAATYTLILNELLTSKNSITREMAENDYVIFPVTPNDGYVYTHTRMRKNRS 118
 QY 207 FYANNHCIGTDLSNPNVSKHMCESGSSSCSEYTCGLYPESEPEVAVASFLRRNINQI 266
 DB 119 PNPDSLCAGTDNRNRW-NHNMEDGIGSSRCPTETYGKKAPESEVERSRSDTLKLGKI 177
 QY 267 KAYISMHSYOHIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKTRTHGSGSETLY 326
 DB 178 KYVLAHFSYSQLLPYPYGHQCHTYNHDDLOAIGDAARSLSAQ-RYGDYTVGINIDALY 236
 QY 327 LAPGGDDWY---DLGIKYSFTSNP-----PVEKLIP 356
 DB 237 PASGSDMAYDTLDIPATYELRPRDGMNGFOLPANDIIP 278

Search completed: September 18, 2002, 16:47:48
 Job time: 280 sec

OM of: US-09-980-881-2 to: GenEmbl: * out_format : pfs

Date: Sep 18, 2002 6:53 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-O=/cgn2_1/USPRO.spool/US09980881/runat_16092002.140128.10635/app-query.fasta.1.1126
-DB=GenEmbl -QPMF=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELLOP=6.000
-DELLEXT=7.000 -START=1 -MATRIX=Diosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09980881@cgn1.1.5320
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPRY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-2
Query length: 360
Database: GenEmbl: *
Database sequences: 1797656
Database length: 187333701
Search time (sec): 4324.800000

Score_list:	Strid	Orig	ZScore	EScore	Len	Documentation
Sequence	gb.pr:AB011969	+ 1884.00	3351.41	2.6e-178	1546	AB011969 Homo sapiens mRNA for
gb.pat:AR086324	+ 1832.00	3360.29	3.1e-173	1272	AR086324 Sequence 1 from patent	
gb.pr:HMWCPX	+ 1832.00	3357.13	4.6e-173	1728	M75106 Human prepro-plasma cat	
gb.pat:116100	+ 1832.00	3357.01	4.7e-173	1749	116100 Sequence 2 from patent	
gb.pr:BC007057	+ 1827.00	3348.25	1.4e-172	1715	I33526 Sequence 2 from patent	
gb.pr:AB021968	+ 1534.50	2726.80	1.7e-143	1421	AB021968 Mus musculus, carboxy	
gb.pr:AF164524	+ 1534.50	2725.81	1.8e-143	1490	AF164524 Mus musculus mRNA for	
gb.pr:AF186188	+ 1528.50	2715.94	6.8e-143	1430	AF186188 Mus musculus thrombin	
gb.pr:AB042598	+ 1502.50	2668.96	2.7e-140	1425	AB042598 Rattus norvegicus carboxy	
gb.pr:AF190274	+ 700.50	1233.25	2.5e-60	1327	AF190274 Bothrops jararaca cat	
gb.pr:SSC133775	+ 646.00	1136.82	6.3e-55	1251	AJ133775 Sus scrofa farraca cat	
gb.pr:HS242486	+ 641.00	1126.82	2.1e-54	1310	AJ242486 Homo sapiens mRNA for	
gb.pr:BC015338	+ 641.00	1125.74	2.4e-54	1454	BC015338 Homo sapiens, clone M	
gb.pat:167698	+ 639.00	1126.70	2.1e-54	1215	167698 Sequence 1 from patent	
gb.pat:AS1896	+ 639.00	1123.61	3.2e-54	1263	AS1896 Sequence 60 from patent	
gb.pat:AR085861	+ 639.00	1123.61	3.2e-54	1263	AR085861 Sequence 11 from patent	
gb.pat:AS1913	+ 631.50	1110.01	1.8e-53	1284	AS1913 Sequence 77 from patent	
gb.pr:AR085875	+ 628.50	1103.44	4.2e-53	1442	AR085875 Sequence 71 from patent	
gb.pr:MS0CARA	+ 628.50	1103.44	4.2e-53	1442	J05118 Mouse mast cell carboxy	
gb.pr:DOG2AP47	+ 627.00	1102.22	5.0e-53	1251	DOG2AP47 Dog mast cell carboxy	
gb.pr:BC012613	+ 625.50	1096.53	1.0e-52	1674	BC012613 Dog mRNA for zymogen	
gb.pat:AR67368	+ 623.00	1089.45	2.6e-52	2154	AR67368 Sequence 124 from patent	
gb.pat:AX333659	+ 621.50	1091.72	1.9e-52	1332	AX333659 Sequence 4168 from patent	
gb.pr:HMWASP	+ 621.50	1091.72	1.9e-52	1332	M81057 Human procarboxypeptid	
gb.pr:RNM67914	+ 621.00	1090.11	2.3e-52	1427	U67914 Rattus norvegicus mast	
gb.pr:HMWCAKMC	+ 619.50	1086.11	3.9e-52	1622	M27717 Human mast cell carboxy	
gb.pr:G28614	+ 619.50	1086.11	3.9e-52	1622	G28614 human S7S SHGC-35784, s	
gb.pat:AX106749	+ 611.50	1073.97	1.9e-51	1311	AX106749 Sequence 9 from patent	
gb.pat:AR072911	+ 604.00	1064.11	6.6e-51	927	AR072911 Sequence 5 from patent	
gb.pr:GCGBPARR	+ 581.00	1019.13	2.1e-48	921	X64539 G. domesticus mRNA for c	
gb.pat:167699	+ 576.00	1014.03	4.1e-48	921	I167699 Sequence 3 from patent	
gb.pr:AX106753	+ 564.00	983.90	1.9e-46	2128	AX106753 Sequence 13 from patent	
gb.pat:AX083139	+ 563.00	983.69	2.0e-46	1826	AX083139 Sequence 42 from patent	
gb.pat:AS1908	+ 561.00	986.32	1.4e-46	999	AS1908 Sequence 72 from patent	
gb.pat:AR085871	+ 561.00	986.32	1.4e-46	999	AR085871 Sequence 67 from patent	
gb.pat:AS1904	+ 561.00	985.78	1.5e-46	1053	AS1904 Sequence 68 from patent	
gb.pr:AR085868	+ 561.00	985.78	1.5e-46	1053	AR085868 Sequence 64 from patent	
gb.pat:AX000417	+ 554.00	973.24	7.6e-46	1053	AX000417 Sequence 27 from patent	

gb.pat:AX000418 - 554.00 973.24 7.6e-46 1053 | AX000418 Sequence 28 from p
gb.pat:AS1917 + 534.00 973.18 7.6e-46 1059 | AS1917 Sequence 81 from Pat
gb.pat:AR085878 + 554.00 973.18 7.6e-46 1059 | AR085878 Sequence 74 from p
gb.pat:AS1921 + 553.00 971.39 9.6e-46 1059 | AS1921 Sequence 85 from Pat
gb.pat:AR085881 + 553.00 971.39 9.6e-46 1059 | AR085881 Sequence 77 from p

seq_name: gb.pr:AB011969

seq_documentation_block:

LOCUS AB011969 1546 bp mRNA linear PRI 02-FEB-2000

DEFINITION Homo sapiens mRNA for carboxypeptidase B-like protein, complete

cds.

ACCESSION AB011969

VERSION AB011969.1 GI:6855463

KEYWORDS carboxypeptidase B-like protein.

SOURCE Homo sapiens cDNA to mRNA.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (sites)

Auton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.

Isolation, molecular cloning, and partial characterization of a

novel carboxypeptidase B from human plasma

J Biol Chem. 266 (32), 21833-21838 (1991)

2 (sites)

Matsumoto,A.

Isolation, molecular cloning, and partial characterization of a

novel carboxypeptidase B from human plasma

Unpublished (1998)

3 (bases 1 to 1546)

Matsumoto,A.

Direct Submission

Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of

Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan

(E-mail:amatemed.kobe-u.ac.jp, Tel:81-78341-7451)

Sequence updated (10-Apr-1998).

location/Qualifiers

1. 1546

/organism="Homo sapiens"

/db_xref="taxon:9606"

20. 1102

/note="carboxypeptidase B-like protein in human brain"

/codon_start=1

/product="carboxypeptidase B-like protein"

/protein_id="BA090475.1"

/db_xref="GI:6855464"

/translation="MKLCSLAVLPVIVLCEQVHPFQFGQVLAALPRISROYQVLN

LMTYVIVLWQPTADILVKKQVHFFVNASDVNKAHLNVSIGPCSVLAVEDLI

QDQISNDVSPRASASYEYHSLNIEIYTERHDMITKIHIGSFLEYEDLI

LKYSGEQAKNAIWDICGHAHEWISPAICFVIGHNRMRNRSEFVANNHIGTDL

NRNFASKHMCCEGASSSCSEYCGYSEPEPVKAVASFLRININQIKAYISMHSYS

QHTVPEYSTRSKSDHEELISVASEVAIEIKSNTRTHGSGETILAFGGDD

WYIDDIKTSFTSNPEVKLLPLSLK

1546

/note="21 A nucleotides"

BASE COUNT 453 a 323 c 297 g 473 t

ORIGIN

polya_site

alignment_scores:

Quality: 1884.00 Length: 360

Ratio: 5.263 Gaps: 0

Percent Similarity: 99.444 Percent Identity: 98.333

alignment_block:

US-09-980-881-2 x AB011969 ..

Align seg 1/1 to: AB011969 from: 1 to: 1546

|||||MetlyslcysrserleuAlaValleuValprolleValleuPhcysGcl 17

|||||

20	ARGAACTTTGCAGCGCTTGAGAGCTCTGTACCCATTCTTCTCTGTGA	69
17	UGLNLHSLVAlPhlealPhlegInSerGlycInValLeuAlaAlaLeuProA	34
70	GGAGCAAGCTTCCTCGCCTTTCAGACTGAGCCCAAGTCTTAGCGCTCTCTTA	119
34	rgThSerArgInValGInValLeuGlnAsnLeuThrTrpThrTyGlu	50
120	GAACGCTTAGGCAAGTTTCAAGTTCTACAGATCTTACTACAACTATAGC	166
51	ILeValLeuTrpGlnProValThrAlaAspLeuILEValLysLysGly	67
170	ATTGTTCTCTGGCAGCCGGTAACAGCTGACTTATGTGTGAAGAAAACA	219
67	nValHISpRePheValAspAlaSerSpsValAspAsnValLysAlaHISL	84
220	AGTCCATTTTTTTGTAAATAGCATCTATGTGCGCAATATGTAAGGCCATT	268
84	euaAsnValSerGlyLeIProCysSerValLeuLeuAlaAspValGluAsp	100
270	TAAATGTGAGCGGAATTCATGACAGTCTCTGCTGGCAGATGTGGAAAT	319
101	LeuILeGInGlnGlnLeSerAsnAspThrValSerProArgAlaSerAl	117
320	CTTATTCACACACAGATTTCACACGACACAGTACGCCCGGACCTTCGC	369
117	aSerTyTrpGluGlnInTyHISerLeuAsnGlnILETySerTrpIleG	134
370	ATCGATAGTGAACAGTATCAGCTACTAATGAATCTATTCTGGATAC	419
134	LupHILeHrLunArgHISpProAspMeLeuThrLysILEHISLleGly	150
420	AATTTATTAACTAGAGGCGCATCTGATATCTTACAAATAATCCACATTGGA	469
151	SerSerPheGluLysTyProLeuTyValLEuLysValSerGlyLysG	167
470	TCTCATTTTGAGAAATACCACATCTATGTTTAAAGTTTCTGGAAAAGA	519
167	uGLInThrAlaLysAsnAlaILETrpLLeAspCysGlyILEHISLAlaArg	184
520	ACAAACACCCAAAATGCCATATGATGTTGACCTGGAAATTCATGACGACAG	569
184	LutTrpILeSerProAlaPheCysLeuTrpPheILEGlyHISAsnArgMet	200
570	AATGATATCTTCCTCTCTTCTGCTGTGGTGTATAGGCCATATATCCAAAG	619
201	TrpArgLysAsnArgSerPheTyAlaAsnAsnHIScysILEGlyHISAs	217
620	TGGAGAAAGAACCGTCTTCTATGCGAACAATTCATTCGATGGAGACAGA	669
217	PLeAsnSerAsnPheValSerLysHISTrpCysGluGluGlyAlaSerS	234
670	CTGTAAATGGAACCTTGGTCTCCAAAACACTGATGAGGAAGGTGCATCCA	719
334	erSerSerCysSerGluThrTyTrpCysGlyLeuTyProGluSerGluPro	250
720	GTTCCTCATGCTCGGAACCTACTGTGACTTATTCCTGACGACAGACCA	769
251	GluValLysAlaValAlaSerPheLeuArgArgAsnILEAsnGlnILETy	267
770	GAATGTGAAGGAGAGTGGCTATGTTCTTGAGAAATAATCCACAGATTAA	819
267	saLEtyrILeSerMetHISerTySerGlnHISLleValPheProTyS	284
820	AGCATACATCGCATGCTTCACTACTCCACAGATATAGTGTTTCCATATT	869
284	eTyTrpArgSerLysSerLysAspHISGluLeuSerLeuValAla	300
870	CCATTACACGAAAGTAAAGCAAAAGCCATGAGAGACTGCTTAGAGCC	919
301	SerGluValValArgAlaILEAspLysThrSerLysAsnThrArgTyThr	317
920	AGTAAAGCAAGTTCGTGATTAAGAAATTAATGTAATAAATACCAAGTATAC	969

```

317 rrlsglyhlglsySercLuthrleuThrlleuAlaproGlyGlyAspa 334
318 |-----|
970 ACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCGGAGGTGGGAGC 1019
334 sptRpiLeuYrAspLeuGlyIleLysTyrSerPheThrSerAsnProPro 350
335 |-----|
1020 ATTGGATCTATGATTGGGCATCAAAATATTCGTTTACATCAAAACCACT 1069
351 ValGluLysLeuLeuProLeuSerLeuLys 360
352 |-----|
1070 GTAGAGAAAGCTTTGGCGGCTGCTCTCAAAA 1099
seq_name: gb_pat:AR086324

LOCUS documentation block:
LOCUS AR086324 1272 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 1 from patent US 5985562.
ACCESSION AR086324
VERSION AR086324.1 GI:10013090
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 1272)
  Morser,M,John and Nagashima,M.
  Method of detecting thrombotic disease risk associated with plasma
  carboxypeptidase B polymorphisms
  Patent: US 5985562-A 1 16-NOV-1999;
  Location/Qualifiers
    source
      1..1272
      /organism="unknown"
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN

Alignment scores:
  Quality: 1832.00      Length: 415
  Ratio: 5.103          Gaps: 3
  Percent Similarity: 86.506      Percent Identity: 85.783

Alignment_block:
US-09-980-881-2 x AR086324 ..

Align seg 1/1 to: AR086324 from: 1 to: 1272

1 MetLysLeuCySserLeuAlaValLeuValProIleValLeuPheCysGI 17
1 ATGAAAGCTTTGCAGCCCTTGCAAGTCCCTTGACCCATTGTTCTTCTTGGA 50
17 uGlnHsValPheAlaPheGlnSerGIyGlnValLeuAlaIalaLeuPro 34
1 |-----|
51 GGAGCATGCTTCGGCGTTTCACAGTGGCCAAATTCTAGCTGCTTCCCTA 100
34 rGThSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGlu 50
101 GAACCTCTAGGCCAACTTCMAAGTTCTACAGAACTTACTACTACACATATGAG 150
51 lleValleuThrGlnProValThrAlaAspLeuIleValLysLysGSI 67
151 ATTGTTCCTGGCAGCCGGTAAACAGCTGACCTTATGTGAAGAAAAACA 200
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHsL 84
201 AGTCATTTTTTTGTAAMTGCATCTGATGTCACAAATGTGAAGGCCATT 250
84 euAsnValSerGIyLleProCySserValLeuLeuAlaAspValGluAsp 100
251 TAAATGTAGCCGAATTCACATGCACTGCTTGCTGGCAGAGCTGGAAGAT 300
101 leuIleGlnGlnGlnIleSerAsnAspPhrValSerProAlaIalaSerAl 117
301 CTATATCAACAGCAGATTTCCAAACAGACAGTCAACCCCGAGGCCCTCCG 350

```


117 aserlyrlyrgruglnlyrhnlsSerleuAsnGluilleTySerPrpIleg 134
135 ATCGACACTATGACAGCTATCCACTCACTAAATGAATCTATTCTTGGATAG 400
134 IuPheileThrgluarhHisProAspMetLeuThrIysIleHisIlegIy 150
401 AATTATTAACAGAGAGGCAATCCGTATGATTGCTTACAAAAATCCACATTGGA 450
151 SerSerPheGluIuLySTyTyProLeuTyValLeuIysValSerGIyISgl 167
451 TCCCATATTGGAGAGTACCACCTCATCTGTTTAAAGGTTCTGGAAAAAGA 500
167 uclnThrAlaIysAsnAlaIleTPrpIleAspCysGIyIleHisAlaArg 184
501 ACAAAACAGCCAAAAATGCCATATGGATTGAGCTGTGGATCCATCCACAG 550
184 IuTPrpIleSerProAlaPheCysLeuTPrpPheIlegIyHis..... 197
551 AATGAGATCTCTCTCTCTTCTTGCTTGCTGTGCATAGGCCATATAACTCA 600
197 197
601 TTCTATGGGATATAGGGCAATATACCAATCTCCAGAGCTTGATATT 650
197 197
651 CTATGTTATGCCGGTGTTAATGTGGACGGTTATGACTACTACATGAAAA 700
198 ..AsnArgMetIrpArgIyIsnArgSerPheTyAlaAsnAsnHisCys 213
701 ACAATTCGATGAGGAGAAAGAACGCTTCTTCTATGGCAACAATCATCTGC 750
214 IlegIyThrAspLeuAsnSerAsnPheValSerIysHisIstrCysGIuGI 230
751 ATCGGAACAGACCTTAATAGGAACCTTCTTCCAAACCTGTGTGAGGA 800
230 uGIyAlaSerSerSerCysSerGIuThrTyCysGIyLeuTyPrpG 247
801 AGTCGATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCTCG 850
247 IuSerGIuPrpGIuValIyAlaValAlaIleSerPheLeuArgArgAsnIle 263
851 AGTCGAACACAGAAATGAGAGGAGGCTGCTTCTTGTAGAGAAATATC 900
264 AsnGIuIleIyAlaTyIleSerMetHisSerTySerGIuHisIleVal 280
901 AACCGAGTTAAAGCCATCAATCCATGCAATTCATATCTCCAGCATATAGT 950
280 IPrpCotySerTyThrThrArgSerIySerIyAspHisGIuLeuAsn 297
951 GTTTCATATTCCTTATACACGAACTAAAGCAAAAGACCATAGAGAACTGT 1000
297 erLeuValAlaSerGIuAlaValAlaIleAspIySThSerIyAsn 313
1001 CTTGATGATACCCAGTGAACACAGTCTGCTGATTGAGAAAACTAGTAAAAAT 1050
314 ThnArgTyThrHisGIyHisGIySerGIuThrLeuThrLeuAlaIleArg 330
1051 ACCAGGATATACCAAGGCCATGGCCACAAACCTTATACCTAGTCTCTGG 1100
330 yGIyGIyAspAspArPrpIleTyAspLeuGIyIleIySTySerPhe... 345
1101 AGGTGGGAGCGATTGGATCTATGATTGGGCATCAAAATATTCGTTTACAA 1150
346Th 346
1151 TTGAACCTTCGAGATACGGGACATACGGGATTCCTTGCTCCGGACGTTAC 1200
346 rSerAsnPrpCovaGIyIyLeuLeuProIleSerIyAsn 360
1201 ATCAAAACCCACTGTAGAGAAAGCTTTTCCGCTGTCTTAAAA 1243

```

seq_name: gb_pf:HMPCPBX

seq_documentation_block:
LOCUS       HMPCPBX                               1728 bp    mRNA    linear    PRI 07-JAN-1995
DEFINITION  Human prepro-plasma carboxypeptidase B mRNA, complete cds.
ACCESSION   M75106
VERSION     M75106.1 GI:189686
KEYWORDS    plasma carboxypeptidase.
SOURCE      Homo sapiens liver cDNA to mRNA.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1728)
AUTHORS     Eaton,D.L., Malloy,B.E., Tsai,S.-P., Henzel,W. and Drayna,D.
TITLE       Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
JOURNAL     J. Biol. Chem. 266 (32), 21833-21838 (1991)
MEDLINE     92042093
FEATURES
    source
        1..1728
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="liver"
    gene
        1..1728
            /gene="PCPB"
        20..1291
            /gene="PCPB"
        /note="prepro-plasma carboxypeptidase B"
        /codon_start=1
        /protein_id="AA06042.1"
        /db_xref="GI:189687"
        /translation="MKLCSLALVLPVILFCEQHFVAFQSGVLAALPRTSRQVLYQNM
            LTTTETVLQPVTAADLIIVKKQVHFVFNASDVNVNAHLNVSIGPSVLADVEDLI
            QOQSLNDPVSFASASYEEOYSLNEIYSIEFTTERHPDLTKIHGSSPEKPLVYV
            LKSGKEQTAANALIMIDGILAREMISPAFCWFIQHTQYGIIGQYTNLRLLVDYV
            TVCLNLYDSGEYVSMKKRMWRKNSFANNHCITGTDLNRPASKHMCEGASSSCEB
            TYCLLYPSESEPEYKAVASFLRNINQIKATISMSTYSGHYFPYSTRSKSDHEELS
            LVASEAVALEIKTSKNRTYTHGSESTLYLAPGGDDMTYDLGIKISFTLELRDTGTVY
            GELLPERYIKPTCREAFPAVASKIAMHVIARN"
    mat_peptide
        86..1288
            /gene="PCPB"
            /product="plasma carboxypeptidase B"
BASE COUNT      518 a      354 c      338 g      518 t
ORIGIN
alignment_scores:
    Quality: 1832.00      Length: 415
    Ratio: 5.103          Gaps: 3
    Percent Similarity: 86.506      Percent Identity: 85.783

alignment_block:
US-09-980-881-2 x HMPCPBX ..
Align seg 1/1 to: HMPCPBX from: 1 to: 1728

1 MetLysLeuCySserLeuAlaValLeuValProIleValLeuPheCysGI 17
20 ATTCAGCCCTTGACGCTTGCGAGCTCTGTAACCCATTGGTCTCTTCGTGTA 69
17 uGlnHISValPheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuPro 34
70 GCAGCATGTCCTTGCGCTTCAGAGTGGCCAGTTCACGCTTCCTCCGA 119
34 rGTHISerArgGlnValGlnValLeuGlnAsnLeuThrThrTYGlu 50
120 GAACCCCTGAGGCAAGTTCAAGTCTCAGAGATCTTACTCAACATATGAG 169
51 ILleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysLysGI 67
170 ATTCGTCCTCGGCGACGCCGGAACAGCTTATTTGTGACAGAAAAAACA 219
67 nValHISphePheValAlaSnAlaSerAspValAspAsnValLysAlaHisL 84

```



```

134 lurbellierhgluarhisprometleuthrlisilegly 150
|||||
441 AATTATACAGAGGACATCTGATATGCTTACAAAAATCCACATTGGA 490
151 SerSerPheglulysylrproleuthrvalleuylserglyscl 167
|||||
491 TCCTCATTTGAGAGTACCCACTGATGTTTAAAGGTTCTGGAAAAAGA 540
167 uglInthralasanaialetprieaspcysglyllehislaarag 184
|||||
541 ACAAGACGCAAAAATGCCATATGATGACTGTGGAAATCCATGCCAGAG 590
184 lurtprileserrproalaphecysleuthrphelleglyhis..... 197
591 AATGATCTCTCCGCTTCTGCTTGCTGATGATGATGATGATGATGATG 640
197 ..... 197
641 TTCTATGGATATATAGGCAATATACCAATCTCTGAGGCTTGATGATTT 690
197 ..... 197
691 CTATGTTATGCCGCTGTTATATGACGGTATGACTACTCATGCAAAA 740
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCys 213
|||||
741 AGAATCGAAATGTGGAGAAAGAACCGTTCTTCTATGCAAAATCATATGC 790
214 lileglythrAspleuAsnSerAsnPhelValSerlyshistrcysglucl 230
|||||
791 ATCGGAACAGACCTGAATAGAGACTTGTCTTCAACACATGCTGTGAGGA 840
230 uglYalaserSerSerSerSerSerSerSerSerSerSerSerSerSer 247
|||||
841 AGGTGATTCAGTCTCTCATGCTCGGAAACCTACTGTGGACTTTATCTGT 890
247 lusergluprogluvallysalavalasaserPheleuargArgAsnIle 263
|||||
891 AGTCAGAACAGAGAGTAGAGCACTGCTACTTCTTGAGAAAGAAATATC 940
264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
|||||
941 AACCAAGATTAAAGCATACATCAGCATGATTCATCTCCAGCATATAGT 990
280 lPheProTyrSerTyrThrArgSerLysSerLysAspHisGluLys 297
|||||
991 GTTTCATATTCTATACAGAACTAAAGCAAGACCATGAGCAACTGT 1040
297 erleuValaserGluAlaValArgAlaIleAspLysThrSerLysAsn 313
|||||
1041 CTCTAGTAGCCAGTAGAGCACTGCTATGTAAGAAACTAGTAAGAAAT 1090
314 ThrArgTyrThrHisGlyHisGlySerGluThrleuThrleuAlaProgl 330
|||||
1091 ACCAGGATATACATGAGCATGCTGCAAAAACCTTATACCTAGCTCCGG 1140
330 yglYalAspAspTrpIleTyrAspLeuGlylleLysTyrSerPhe... 345
|||||
1141 AGGTGGGAGCATTTGATCTATGATTTGGGATCAAAATATTCGTTTCAA 1190
346 .....Th 346
1191 TTGAACCTTCGATACGGGACATACGATTCCTGCTGCCGAGCGTTAC 1240
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
|||||
1241 ATCAACCCACCTCTAGAGAGCTTTGCCCCCTGCTCTTAAAA 1283
seq_name: gb_pat: I33526
seq_documentation_block:
LOCUS I33526 1749 bp DNA linear PAT 06-FEB-1997

```

```

DEFINITION Sequence 2 from patent US 5593674.
ACCESSION I33526
VERSION I33526.1 GI:1824317
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.T. and Eaton,D.L.
TITLE Plasma carboxypeptidase
JOURNAL Patent: US 5593674-A 2 14-JAN-1997;
FEATURES
Source
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN
alignment_scores:
Quality: 1832.00 Length: 415
Ratio: 5.103 Gaps: 3
Percent Similarity: 86.506 Percent Identity: 85.783
alignment_block:
US-09-980-881-2 x I33526 ..
Align seg 1/1 to: I33526 from: 1 to: 1749
1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGly 17
|||||
41 ATGAAGCTTTGCAGGCTTGACAGTCCCTTGATCCCATTTGCTCTGTGGA 90
17 uglInthralPhealaphelinserglyGlnValleuAlaalaProa 34
|||||
91 GCAGATGCTCTCGCGTTTCAGAGTGGCCAGTTCTAGCTGCTCTCTGA 140
34 rgnrSerArgGlnValGlnValleuGlnAsnLeuThrThrTyrGlu 50
|||||
141 GAACTCTAGGCAAGTCAAGTTCACAGAAATCTTACTACCAACATATGAG 190
51 lleValleuTrpGlnProValThrAlaAspLeuIleValLysLysGly 67
|||||
191 ATTGTTCTTGCGACCGGTTAACAGCTGACCTTATTTGAGCAAAAACA 240
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
|||||
241 AGTCATATTTTGTAAATGATCTGATGTGACAAATGGAAGCCATT 290
84 euAsnValSerGlyLleProCysSerValLeuLeuAlaAspValGluAsp 100
|||||
291 TAAATGTAGCGGAATTCATGCACTGCTGCTGCGAGACGTGGAGAT 340
101 leuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
341 CTATTCACACAGCAGATTCCACAGCACAGACAGCCCCCGAGCCCTCGC 390
117 aSerTyrGlyGlnGlnTyrHisSerLeuAsnGluLleTyrSerTrpIleG 134
|||||
391 ATCGATCATATGACAGTATCACTCACTAAATGAATCATTTCTTGATAG 440
134 lurbellierhgluarhisprometleuthrlisilegly 150
|||||
441 AATTATATACAGAGGACATCTGATATGCTTACAAAAATCCACATTGGA 490
151 SerSerPheglulysylrproleuthrvalleuylserglyscl 167
|||||
491 TCCTCATTTGAGAGTACCCACTGATGTTTAAAGGTTCTGGAAAAAGA 540
167 uglInthralasanaialetprieaspcysglyllehislaarag 184
|||||
541 ACAAGACGCAAAAATGCCATATGATGACTGTGGAAATCCATGCCAGAG 590
184 lurtprileserrproalaphecysleuthrphelleglyhis..... 197

```

```

|||||
591 AATGATCTCTGCTTCTGCTTGTGCTCATAGGCGCATATACTCA 640
197 ..... 197
641 TTCTATGGATATAGGCAATATACCAATCTCTGAGGCTTGATTT 690
197 ..... 197
691 CTATGTTATGCCGGTGTATATGTGACGGTATGACTACTCATGAAAA 740
198 .. AanaTMeTTPaRgLyAsnAaRgSerPheTyAlaAsnAaHicys 213
|||||
741 AGAATCGAATGTGGAGAAAGAACCGTCTTCTATGTGCAACAAATCATG 790
214 TleGlyThraSPLeuAaSerAsnPheAlaSerLysh1strPCysglu1 230
|||||
791 ATCGAAGACAGACCTGATAGGAATCTTGCTCAAAACACTGGTGTAGGA 840
230 uGlyAlaSerSerSerSerSerSerSerSerSerSerSerSerSer 247
|||||
841 AGGTGCTCAGTCCATGCTCATGCTCGAACCCTACTGCTTATCTCTG 890
247 luserGluPProGluVallyAlaValAlaSerPheLeuArgArgAaHle 263
|||||
891 AGTCAGAACCAAGAGTAAGGCACTGCTAGTCTTCTGAGAGAAATATC 940
264 AaenGlnleuLyAlaTyrlleSerMetH1SerTySerGlnH1sileva 280
|||||
941 AACAGATTAACATACATACATGATCATCTACTCCAGCATATAGT 990
280 lPheProTySerTyThraSerSerSerSerSerSerSerSerSerSer 297
991 GTTTCATATCTCTATACAGAAAGTAAAGCAAGCAAGCAAGCAAGT 1040
297 eRLeuValaSerGluAlaValAlaAlaSerPheSerSerSerSer 313
|||||
1041 CTCTAGTAGCAGTGAAGCACTGCTGCTATGAGAAACTACTATAAAAT 1090
314 ThraTyThraH1sGlySerGluThraLeuTyLeuAlaProG1 330
|||||
1091 ACCAGATTAACATGATGCTGCTGAGAAACCTTACTACTCTCTG 1140
330 yGlyGlyAaPhePhePhePhePhePhePhePhePhePhePhePhe 345
|||||
1141 AGGTGGGAGCATGATGATGATGATGATGATGATGATGATGATGAT 1190
346 .....Th 346
1191 TTGAACTTGAGATACGGGACATACGATCTTGCTGCGGAGCGTTAC 1240
346 rSerAaPProProValGluLyLeuLeuProLeuSerLeuLyS 360
|||||
1241 ATCAAAACCACTGTAGAGAGCTTTGCCGTGCTCTTAAAA 1283

seq_name: gb.Pr:BC007057

seq_documentation_block: 1715 bp mRNA linear PRI 12-JUL-2001
LOCUS BC007057
DEFINITION Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
IMAGE:3934520, mRNA, complete cds.
ACCESSION BC007057
VERSION BC007057.1 GI:13937896
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer

```

REMARK

COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapps-r@mail.nih.gov
 Tissue Procurement: CLOMTECH
 cDNA Library Preparation: CLOMTECH Laboratories, Inc.
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mdickpax1@stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Series: IRAL Plate: 16 Row: f Column: 21
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4503004.

Location/Qualifiers

```

1..1715
/organism="Homo sapiens"
/db_xref="locusID:1361"
/db_xref="taxon:9606"
/clone="MGC:12495 IMAGE:3934520"
/tissue.type="Skeletal Muscle"
/clone_lib="NIH MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
25..1296
/codon_start=1
/product="carboxypeptidase B2 (plasma)"
/protein_id="AA07057.1"
/db_xref="GI:13937897"
/translacion="MKLSLAVALPVLVFEQHYFAFQSGQVLAALPRTSRQVQLQ
LTTYTELVLPQVPAADLVKRVQVHFNASDVDAVNAHNVGIPCSYLAIVEDLI
QOOISNDVSPRASAYEQQHISNIEYSWIEFTEHHPMLTKIHIGSFEEKYPLY
LKVSGEQAARKNAIMIDCGIHAEMISPARCMEFIGHITQFVIGIOYTLALNPFY
VMPVAVNDGVDYSWKRMKRMKNSPFYANNHCIGTDLRNRFPAKHNCEBASSSSE
TYGGLVPESEPEKAVASPLKRNINQKATISMASHSDHIVPFYSTRSKROHEELS
LVASENRALEKSKNRTYTHGHSESLYLAPEGGDWDVLDGIRKYSFTLELDGTY
GFLPEKTYIPTRKEAFNAVSKIAHMYHVRV"

```

CDS

```

BASE COUNT      518 a      348 c      341 g      508 t
ORIGIN
alignment_scores:
  Quality: 1827.00      Length: 415
  Ratio: 5.089      Gaps: 3
  Percent Similarity: 86.506      Percent Identity: 85.542
alignment_block:
US-09-980-881-2 x BC007057 ..
Align seg 1/1 to: BC007057 from: 1 to: 1715

```

```

1 MetLyLeuCySerLeuAlaValLeuValProLeuValLeuPheCyG1 17
|||||
25 ATGAGCTTTGCGAGCTTGCAGTCTTGTACCAATGTTCTCTGTGGA 74
17 uGlnH1sValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
|||||
75 GCAGCATGCTTCGGCTTACAGAGTGCAGAGTCTACTCTCTCTCTTA 124
34 rGThSerArgGlnValGlnValLeuGlnAsnLeuThThThTyGlu 50
|||||
125 GAACCTTAGCGCAAGTTAAATTTACAGAAATCTTACACAACTATGAG 174
|||||
51 TleValLeuTyGlnProValThraAlaSPLeuLeuVallySLySLyG1 67
|||||
175 ATTGTTCTCTGCGAGCGGTAAAGCTGACTTATTTGTGAAGAAAAACA 224

```

```

67 nValHisPheValAlaSerAspValAlaSerValLysAlaHisL 84
68 AGTTCATTTTGTAAATGATCTGATGCGAAGTGTGAAACCAT 274
225 eValSerGlyIleProCysSerValLeuValAlaSerValAla 100
226 TAAATGAGAGGAAATTCATGAGTCTTTCGCGAGAGAGTGAAG 324
275 TAAATGAGAGGAAATTCATGAGTCTTTCGCGAGAGAGTGAAG 324
325 CTATATCAACAGAGATTTCCAAAGCAGACAGTCCCGGAGCTCG 374
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAla 117
102 TAAATGAGAGGAAATTCATGAGTCTTTCGCGAGAGAGTGAAG 324
325 CTATATCAACAGAGATTTCCAAAGCAGACAGTCCCGGAGCTCG 374
117 aSerTyTyGlnGlnIleTyHisSerLeuAsnGlnIleTySer 134
118 ATGCTACTATGAAACAGTATCTCATTAATGAAATCTATCTTGA 424
375 ATGCTACTATGAAACAGTATCTCATTAATGAAATCTATCTTGA 424
134 IupheIleThrGluArgHisProAspMetLeuThrLysIleHis 150
135 ATTTTATTAATGAGAGATCTGATATGCTTACAAAATTCACATT 474
425 ATTTTATTAATGAGAGATCTGATATGCTTACAAAATTCACATT 474
151 SerSerPheGluLysTyTyProLeuTyValLeuLysValSer 167
152 TCCATATTTGAGAGTACCCACTGATGTTTAAAGTTCTGAGAAA 524
475 TCCATATTTGAGAGTACCCACTGATGTTTAAAGTTCTGAGAAA 524
167 uGlnThrAlaLysAsnAlaIleTyrPheAspCysGlyTyrHis 184
168 ACAAGCGCCAAATAATCCATATGATGATGCTGGAATCATCCAG 574
525 ACAAGCGCCAAATAATCCATATGATGATGCTGGAATCATCCAG 574
184 IupPheIleSerProAlaPheCysLeuTyrPheIleGlyHis 197
185 AATGATCTCTCCGCTTCTGCTTGCTTGCTGATGAGCATATAC 624
575 AATGATCTCTCCGCTTCTGCTTGCTTGCTGATGAGCATATAC 624
197 ..... 197
625 TTCTATGGATTAATAGGCAATATACCAATCTCTGAGGCTTGAT 674
197 ..... 197
675 CTATGTTATGCCGCTGTTAATGATGATGTTATGACTACTCAT 724
198 .AsnArgMetTyrPargLysAsnArgSerPheTyrAlaAsnHis 213
199 AGAATCCAAATGAGAGAAAGACCGTCTTCTTATGCGAACAATCA 774
725 AGAATCCAAATGAGAGAAAGACCGTCTTCTTATGCGAACAATCA 774
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTyr 230
215 ATCGAAGACAGCTGATAGGAATCTTTCGTTCCAAACAGCTGTG 824
775 ATCGAAGACAGCTGATAGGAATCTTTCGTTCCAAACAGCTGTG 824
230 uGlyAlaSerSerSerSerSerGluThrTyrCysGlyLeuTyr 247
231 AGGTGCATCCAGTCTCATGCTCGAAGACCTACTGTTGATTTAT 874
825 AGGTGCATCCAGTCTCATGCTCGAAGACCTACTGTTGATTTAT 874
247 IuSerGluProGluValLysAlaValAlaSerPheLeuArgArg 263
248 AGTTCAGAACGAGAGTGAAGGACAGTGTGTTCTTGAGAGAAAT 924
875 AGTTCAGAACGAGAGTGAAGGACAGTGTGTTCTTGAGAGAAAT 924
264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHis 280
265 AACGAGATTAAAGCATATCATGATCATATCCACGATATAGT 974
925 AACGAGATTAAAGCATATCATGATCATATCCACGATATAGT 974
280 IuPheProTySerTyThrArgSerTySerLysAspHisGlnL 297
281 GTTTCCTAATTCCTATACCAAGTAAAGCAAGACATGAGCACT 1024
975 GTTTCCTAATTCCTATACCAAGTAAAGCAAGACATGAGCACT 1024
297 eLeuValAlaSerGluAlaValAlaArgAlaIleAspLysThr 313
298 CTCTAGAGCCAGTGAAGCACTTCTGCTATTTGAGAAACCTAGT 1074
1025 CTCTAGAGCCAGTGAAGCACTTCTGCTATTTGAGAAACCTAGT 1074
314 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyrLeuAla 330
1075 ACCAGGATATACATGGCAGTGGCTCAGAAACCTATATACCTAG 1124
330 yGlyGlyAspAspTyrPheTyraPheLeuGlyIleLysTySer 345

```

```

1125 AGGTGGGAGCATTTGATATGATTGGCATCAAAATTCGTTTACA 1174
346 .....Th 346
1175 TTGAATCTCGAATACGGGACATACGATCTTCTCCGAGCGTTAC 1224
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
1225 ATCAAAACCCACTGTAGAGAGCTTTTCCGCTGCTCTAATAA 1267

seq_name: gb_ro:AB021968
seq_documentation_block:
LOCUS AB021968 1421 bp mRNA linear ROD 27-JUL-2000
DEFINITION Mus musculus mRNA for carboxypeptidase R, complete cds.
ACCESSION AB021968
VERSION AB021968.1 GI:9558447
KEYWORDS carboxypeptidase R.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Sato,T., Miwa,T., Akatsu,H., Matsukawa,N., Ohta,K., Okada,N.,
Campbell,W. and Okada,H.
TITLE Pro-carboxypeptidase R is an acute phase protein in the mouse,
whereas carboxypeptidase N is not
JOURNAL J. Immunol. 165 (2), 1053-1058 (2000)
MEDLINE 20341711
REFERENCE Sato,T.
AUTHORS 2 (bases 1 to 1421)
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1998) Tomoo Sato, Fukushima Hospital, Choji
Medical Institute, 19-14, Azayamanka, Noyori-cho, Toyohashi, Aichi
441-8124, Japan (E-mail:tomoo@cep-ip.or.jp, Tel:81-532-46-7511,
Fax:81-532-46-4895)
FEATURES
source location/Qualifiers
1..1421
/organism="Mus musculus"
/db_xref="taxon:10090"
17..1285
/product="carboxypeptidase R"
/codon_start=1
/protein_id="BAB03402.1"
/db_xref="GI:9558448"
/translation="MKLHGLGILVATILYEOHGFAROSGOVLSAPRTSRQVLLQNL
TTTYEVVLMQPTAETIEKKEVHEFVNASDVSKAHLNVSRIPEVNLNNVEDLIE
QOEFNDVSPRASASYEQYHSINLETYSMEVTEQHPMLQKITYIGSSFEKPYVL
KVSGEORIKNAIWDICGIHAREMISPAFCIMFIVGTFQHGKENDYTRILRRVDYI
MPMANVDGYDWTWKKNRMWRKNSAKNNRCVGTIDRNPFASHMCSEKSSGSET
YGLYSEPEYVAADPLRRNIDHKAYISMYSQOILFPYSYRSKSKDHEEEL
VASEAFAIESIKNRYTHGSSSESLYLAPEGSDMIYDILGIKYFTTLELRDTGRYG
FLPERIKPTCEALALAIISKIYWHVIRNT"
1421
polyA_site /note="11 a nucleotides"
BASE COUNT 418 a 320 c 305 g 378 t
ORIGIN
alignment_scores:
Quality: 1534.50 Length: 415
Ratio: 4.540 Gaps: 4
Percent Similarity: 81.446 Percent Identity: 72.530
alignment_block:
US-09-980-881-2 x AB021968 ..
Align seg 1/1 to: AB021968 from: 1 to: 1421
1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysG 17
|||||
17 ATGAAGCTTCATGCGCTTGGAATCTGTTAGCAATCATCTCTAT...GA 63

```

```

17 ucghhishvalrphelapheglnserglylvalleualaalaaleuproa 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
64 GCACCATGGCTTCGCTTCAGAGTGGCCAGGTTTATCTCTCTCCAA 113
34 rgttrserarglnvalglnvalleuasnleuthrthrtthrtgln 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
114 GAACCTCAGGCAAGTTCACTTCAAGATCTGATCTACACAGCTATGAG 163
51 lvalleuthrpglnprovalthralsprrleuallvalylslysgl 67
::: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
164 GTCCTTCTCGGAGCCAGTGCAGCTGAATTCATTCGAGAAAGAAAGCA 213
67 nvalhishrpherhvalasnalsasrvalasprvalylslysl 84
::: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
214 AGTCACATTTTGTGTGAATGCGTGTGATGTCAGACAGTCAACCGCAT 263
84 euasnvalserglyllepocyservalleuualaspvalglnasp 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
264 TAAATGACAGACAAATTCATTTACCTTCTGATGAACAACGCTGGAGAC 313
101 leuileglnlgnlleserasnspthrvalserproargalaserai 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
314 CTATATGACAGACAGACTTCAATGACACAGCTGACCCCGCGCTCCG 363
117 asertrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
364 TTCACTACTATGAGAGATATCATCTGGTAAATGAAATCTATCTGTGATG 413
134 lupheilethrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 150
|| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
414 AAGTCATTAAGTGAACAGACATCTGACATGCTCCAGAAATCTACATCGCA 463
151 serSerPheglulysrtyrtyrtyrtyrtyrtyrtyrtyrtyrtyr 167
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
464 TCAATTCATTCAGAAAGTACCCTTATGTTTAAAGTCTCAGAGAAAGA 513
167 ugnlthrthralysasnlatlertprileaspgsglylethialargc 184
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
514 ACAAGAAATCAAAATGCCATCTGATCGATCGATGAAATCCATGACGAG 563
184 lurtprileserproalrphesysleuthrphelleglnhis 197
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
564 AATGATTTTACCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 613
197 ..... 197
614 TTCATGGGAAAGAAATCTGTATACAGACTTGTAGGACGCTGATTT 663
197 ..... 197
664 CTACATCATGCCCGTGTATGAGTGTATGGCTATAGCTACAGCTGAAAA 713
198 .. AsnArgheTtrPratgLyAsnArgSerPheThrAlaAsnHisCys 213
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
714 AGAATCGAATGTGGAGAAACCGCTCTGCTACAAAGCAACACCGCTCG 763
214 lilegltThrAspLeuAsnSerAsnPhelValSerLysHisTrpCysGlu 230
::: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
764 GTGGGACAGACCTGACACAGAACTGCTCCAAACATGCTGTGTAGAA 813
230 uGlyAlaSerSerSerCysSerGluThrTyrcysGlyLeuTyrrpG 247
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
814 AGTTCGCTCAAGTCTCTCTCTGAAACCTACGTGTGACTTATCTCG 863
247 luserGluProGluValLysAlaValAlaSerPheLeuArgArgAsnile 263
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
864 AATCTGAGCCAGAGGTGAAGGAGTGGCTGACTCTTGAGAGAAATATC 913
264 AsnglnlileValAlaTyrlleSerMetHisSerTyrsGlnHisIleVa 280
::: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
914 GACCAATTAAGCTTACATCTAGTATGACATCATCTCCCAACAAATACT 963

```

```

280 lPheProTyrSerTyrThrArgSerLysSerLysAspHisGlnGluLeu 297
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
964 GTTTCCTATTCCTATATACAGAACAAAGCAGACAGCAAGACTGT 1013
297 erLeuValAlaSerGluAlaValArgAlaIleAspLysTrpSerLysAsn 313
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1014 CTCTAGTGGCCAGCAGACAGATTCGTCAATTGAAATATTAATAAAC 1063
314 ThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrlleuAlaProG 330
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1064 ACCAGGTACACACAGGAGCTGCTCAGAAAGTTATATCTACTCTCTG 1113
330 yGlyGlyAspAspTrpIleTyrrAspLeuGlyIleLysTyrSerPhe... 345
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1114 AGGTTTGCAGATTTGATCTATGATTTGGGCATCAATATATCTTTACAA 1163
346 .....Th 346
1164 TTGAGCTCCGAGATACAGGCAGATACGAGATTCCTGCTCAGAGATAC 1213
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1214 ATCAAAACCCACTTGTGACAGAAAGCTTTGGCCGACATCTTAAAA 1256
seq_name: gb_ro:AF164524
seq_documentation_block: 1490 bp mRNA linear ROD 05-APR-2000
LOCUS AF164524
DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (tafi)
ACCESSION AF164524
VERSION AF164524.1 GI:7416966
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE
AUTHORS Marx,P.F., Wagenaar,G.T.M., Reijerkerk,A., Tlekstra,M.J., van Rossum,A.G.S.H., Gebbink,M.F.G.B. and Meijers,J.C.M.
TITLE Characterization of mouse thrombin-activatable fibrinolysis inhibitor
JOURNAL Thromb. Haemost. 83 (2), 297-303 (2000)
MEDLINE 20201996
REFERENCES
AUTHORS Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1999) Haematology, DMCU, Heidelberglaan 100, Utrecht 3584 CX, The Netherlands
FEATURES
SOURCE
1. 1490
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:477035"
1. 1490
/gene="Tafi"
84. 1352
/gene="Tafi"
/function="carboxypeptidase"
/codon_start=1
/product="thrombin-activatable fibrinolysis inhibitor"
/protein_id="AF62385.1"
/db_xref="GI:7416967"
/translation="MKLHGIGLIVAILIYEQHFAROSGVLSALPRTSROYOLLNT
TTTVEVLMQPTAEFELEKKEVFEVNASDQSVAAHINVRIPNTVMNVEDLIE
OOFNDVSPRASASYROYHINLFTSMVLEIVTEBHPMLDKIYGSSEFYPLVL
KVGSKROIKMAIWDCCGHARWISDPACLMFTGVTQFHKENLYTLRLHVDVYI
MPWANDGDIYTKKNRMKNSAHKNNRCVGTDLNRFASKHMCENKASSSCSET
YCGLYPESSEPVAAVDFLRNIDIKAVISMHSYQQILFFYSYRKSXKHDELST
VSEAVRAIESINKNRYTHGSGSESLYLAPEGSDMWYDLGIKSYFTLELDTGRYG
FLPERYIRPTCEALAAISKIYWHVIRNT"
BASE COUNT 438 a 337 c 322 g 393 t
ORIGIN

```



```

/protein_id="AF00528.1"
/db_xref="GI:6003652"
/translation="MKLHGIGLIVAILLEQHGFAFGOSVLSALPRTSNOVOLLNL
TTTTVEVLQMPYAEFEKKKEVHFEVNASDVSVSAHLNVSIIPNVIMNVEDLIE
OOFENDVSRBASYYEBOYHSLNETYSMEVTEBOHPMLQKITYIGSSPERKPIYL
KVGKROBKIKMNIWDCGSHAREMISPARCLMTIGVYTOHGENLYTLRLRVDTYI
MPVMNDGIDYTYKRRMRKRNKSNKSVGTDLNPFASHHWCERKSSCSSET
YCOLIPSEEVAAVADFLRNIDHIKAYISMHSQOILFPYSYNSRSKSDHEELSL
VASEAVRAIGSEFNKNTRYTHGSGESLYLAPGSDWDIMYDLGIKYSFTIELRDYGRY
FLPERRYIKPTCAELAAISKIYMHVIRMT"
BASE COUNT      421 a      320 c      310 g      379 t
ORIGIN

alignment_scores:
  Quality: 1528.50      Length: 415
  Ratio: 4.536          Gaps: 4
Percent Similarity: 81.205      Percent Identity: 72.289

alignment block:
US-09-980-881-2 x AF186188 ..
Align seg 1/1 to: AF186188 from: 1 to: 1430

1 MetLysLeuCySerLeuAlaValLeuValProLeuValLeuPheCysG1 17
|||||.....|.....|.....|.....|.....|.....|
20 ATGAAGCTTCATGGCTTGGATCTGTGAGCCCTCATCTCTAT...GA 66
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaValLeuProa 34
|||||.....|.....|.....|.....|.....|.....|
67 GCACATGGCTGGCTTGCCTTGAGAGTGGCCAGTTTATCTGCTTCCAA 116
34 rGmTrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrGlu 50
|||||.....|.....|.....|.....|.....|.....|
117 GAACCTCCAGGCAAGTCAACTACTCTCAGATCTTACTACACCTATGAG 166
51 lLeValLeuTrpGlnProValThrAlaSerLeuLeuValLysLysG1 67
:::|.....|.....|.....|.....|.....|.....|
167 GTGCTTCTCGGACGACGAGTGCAGCGCTGAATTCATCGAGAAAGAA 216
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
:|.....|.....|.....|.....|.....|.....|
217 AGTCCACTTTTGTGTGAAATGCTGTGATGTCGACAGTGCAGGCGCAT 266
84 euAsnValSerGlyLleProCySerValLeuLeuAlaAspValGluasp 100
|||||.....|.....|.....|.....|.....|.....|
267 TAAATGTGACGACAGATTCATTTAACGTCCTGATGAACAACGCGAGGAC 316
101 LeuIlleGlnGlnLlleSerAsnAspThrValSerProArgAlaSerAl 117
|||||.....|.....|.....|.....|.....|.....|
317 CTAATTGAACAGCAGACTTCATATGACAGCGTACGCCGCCGCTCCGC 366
117 aSerTyTyTyGlnGlnTyHisSerLeuAsnGlnLleTySerTrpLleG 134
|||||.....|.....|.....|.....|.....|.....|
367 TTATATCTATGACAGTATCATCTGCTAAATGAATCTATCTGATGATG 416
134 luePheThrGlnArgHisProAspMetLeuThrLysLleHisLleGly 150
||.....|.....|.....|.....|.....|.....|
417 AAGCTTTAACTGACACGATCTGACATGCTCCAGAAATCTAATCGGA 466
151 SerSerPheGlnLysTyTyProLeuTyValLeuLysValSerGlyLysG1 167
|||||.....|.....|.....|.....|.....|.....|
467 TCATCATTCGAGAGTACCCACTTATGTTTAAAGTCTCAGAGAAAGGA 516
167 uGlnThrAlaLysAsnAlaLleTrpLleAspCysGlyLleHisAlaArg 184
|||||.....|.....|.....|.....|.....|.....|
517 ACMAAGAAATCAAAAATGCATCTGATCGACTGGAATCCATGCGACAG 566
184 lueTrpLleSerProAlaPheCysLeuTrpPheLleGlyHis..... 197
|||||.....|.....|.....|.....|.....|.....|
567 AATGCATTTACCTGCTTCTGTTGTGTGATGCTAAGCTGACGACACA 616
197 ..... 197

```

```

617 TTCCATGGGAAAGAAANTCTGTATACAGACTTCTGAGGACCGTGATTT 666
197 ..... 197
667 CTACATCATGCCCGCTGATGAACGTGATGCTATGACTACACGTGAAA 716
198 ..AsaArgMetTrpArgLysAsnArgSerPheTyAlaAsnAsnHisCys 213
|||||.....|.....|.....|.....|.....|.....|
717 AGAATCCAAATGTGAGAGAGAACCGCTGCTGCTACAGAAACACCGCTGC 766
214 lLeGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGlu 230
:::|.....|.....|.....|.....|.....|.....|
767 GTGGCAGACAGCTGAAACAGAACTTGCCTCCAAACAGTGTGTGAGA 816
230 uGlyAlaSerSerSerSerCysSerGluThrTyTyCysGlyLeuTyPro 247
|||||.....|.....|.....|.....|.....|.....|
817 AGTGCGTCAAGTCCCTGCTGCTGAAACCTACTGCTGATGATTCCTG 866
247 lueSerGluProGluValLysAlaValAlaSerPheLeuArgAsnLle 263
|||||.....|.....|.....|.....|.....|.....|
867 AGTCTGAGCCAGAGGTGAAGCGAGTGGCTGACTTCTTGAGAAAGAAAT 916
264 AsnGlnLleLysAlaTyTrlSerMetHisSerTySerGlnHisLysLe 280
:::|.....|.....|.....|.....|.....|.....|
917 GACCACATTAAGCTTACATCAGATGACACTCATCTCCCAAAATATCT 966
280 lPheProTySerTyThrArgSerLysSerLysAspHisGluGluLeu 297
GTTTCCCTATTTCTTATPACAGAAAGCAAAAGCAAGCAAGCAAGCAAG 1016
297 eLeuValAlaSerGluAlaValArgAlaLleAspLysTrSerLysAsn 313
|||||.....|.....|.....|.....|.....|.....|
1017 CTCTAGAGGCGACGAGCAAGCGTTCGCAATTTGAAGTTTATTAATAAC 1066
314 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyLeuAlaProG1 330
|||||.....|.....|.....|.....|.....|.....|
1067 ACCAGGTACACACACGCGAGTGGCTCAGAAAGTTTATCTAGCTCTCG 1116
330 yGlyGlyAspAspTrpLleTyTraspleuGlyLleLysTySerPhe... 345
|||||.....|.....|.....|.....|.....|.....|
1117 AGSTTCTGACAGATGATCTATGATTTGGGCACTCAAAATTTGTTGACA 1166
346 ..... 1166
1167 TTGAGCTCCGAGATACAGCAGATACGATTCCTGCTGCTGAGAGATTC 1216
346 lSerAsnProProValGlnLysLeuLeuProLeuSerLeuLys 360
|||||.....|.....|.....|.....|.....|.....|
1217 ATCAAAACCCACTTGTGCAAGACTTTGGCCGCCATCTCTAATAA 1259
seq_name: gb_to:AB042598
seq documentation block:
LOCUS      AB042598      1425 bp      mRNA      linear      ROD 02-DEC-2000
DEFINITION Rattus norvegicus mRNA for pre-procarboxypeptidase R, complete cds.
ACCESSION  AB042598
VERSION    AB042598.1 GI:11526576
KEYWORDS
SOURCE
ORGANISM  Rattus norvegicus CDNA to mRNA.
            Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Ratus.
REFERENCE  1 (sites)
            Kato,T., Akatsu,H., Sato,T., Matsuo,S., Yamamoto,T., Campbell,W.,
            Hotta,N., Okada,N. and Okada,H.
            Molecular cloning and partial characterisation of rat
            procarboxypeptidase R and carboxypeptidase N
            Microbiol. Immunol. 44 (8), 719-728 (2000)
REFERENCE  2 (bases 1 to 1425)
            Kato,T.
AUTHORS

```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Lepisosteidae; Squamata; Scleroglossa; Serpentes; Colubroidae; Viperidae; Crocodylia; Bothrops.

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 1327)
Screening of Bothrops jararaca pancreas cDNA library
Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 1327)
Screening of Bothrops jararaca pancreas cDNA library
Unpublished

FEATUERS
source
1.1327
/organism="Bothrops jararaca"
/db_xref="taxon:8724"
21.1271
/product="carboxypeptidase homolog"
/protein_id="AA01344.1"
/translation="MPEFLIGATSAFAETTVHREDEKRYRTPRNEDEYFLNYL
/db_xref="GI:6013463"
/translation="MPEFLIGATSAFAETTVHREDEKRYRTPRNEDEYFLNYL
ANIVOVDEWRPDSVELVKAEMTVDFRIDEARCSYESEILOSGLNYEILIDLOAVLD
ROLDNHARTAGYNEKYNSMEKIDAMTADIANENPISVRLQIGTFEGRPBLKVG
KPGVKKATFIDCGFARHMSIPACOMPVRAVPTCKETITPOLINKDFTIPLV
NIDGVYSSKOSRMKRSVAGSTCTGDPNRPDMACSVGASRNCSTYCGSK
PSEKETRLADFTIRNSITDAILHISYQMLPYSTYDLSNNKLSIMKEA
INELVAFETETYPGAATTPYPAAGSDMAYDQIGVAFTEFLDRGRYGFALPES
QIKPTCEETMVAKKIAYEMLSLY"

BASE COUNT 416 a 245 c 274 g 392 t
ORIGIN

alignment_scores:
Quality: 700.50 Length: 369
Ratio: 2.780 Gaps: 5
Percent Similarity: 68.293 Percent Identity: 38.482

alignment_block:
US-09-980-881-2 x AF190274 ..

Align seg 1/1 to: AF190274 from: 1 to: 1327

```

15 PheCysGluGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaI 31
   |||:||||| ||| |||:||||| |||:|||||
60 TTTCGTAACAAACAGCTTCACAGATTGATGGGAAAGGTTTACCGCTG 109
   : |||:||||| : |||:||||| : |||:|||||
31 AleuProAlaGlyHisSerArgGlnValGlnValLeuGlnAsnLeuThrThr 48
   : |||:||||| : |||:||||| : |||:|||||
110 TACACCTCGCAATGAGAGTATATTTTCTCAATTACTTACGCCAATA 159
   : |||:||||| : |||:||||| : |||:|||||
48 hCtyrGluLeuValLeuThrGlnProValThrAlaAspLeuValLys 64
   : |||:||||| : |||:||||| : |||:|||||
160 TTCTGCAAGCTGACTTTTGGCCACGATTCGTTGAACTAGTGAAGAAGA 209
   : |||:||||| : |||:||||| : |||:|||||
65 LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys 81
   : |||:||||| : |||:||||| : |||:|||||
210 GAATGACGTGTGATTTCCGAATTTGAGACAGACAGATGTTCTGAAGTTGA 259
   : |||:||||| : |||:||||| : |||:|||||
81 AlaLeuHisLeuAsnValSerGlyLysProCysSerValLeuLeuAlaAspV 98
   : |||:||||| : |||:||||| : |||:|||||
260 GAGCATTTCTACCAAGACAGTGAATTTATTAATTTGATTAATAC 309
   : |||:||||| : |||:||||| : |||:|||||
98 aGluAspLeuLeuGlnGlnGlnHisSerAspAspThrValSerProArg 114
   : |||:||||| : |||:||||| : |||:|||||
310 TTCACAGCTCTCTTGTATGACCACTGGACAACTACCGCT.....CGG 350
   : |||:||||| : |||:||||| : |||:|||||
115 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLysLys 131
   : |||:||||| : |||:||||| : |||:|||||
351 ACCGCTGATACATATGAAATAATATATAGTTGGAAAGATAGAGGC 400
   : |||:||||| : |||:||||| : |||:|||||
131 rTripleGluPheLeuThrGluArgHisProAspMetLeuThrLysIleH 148
   : |||:||||| : |||:||||| : |||:|||||

```

```

401 TTGACACCTGATATTCGCATGAAATCCAAATGCTTTGTTCCGATTAC 450
148 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
   : |||:||||| : |||:||||| : |||:|||||
451 AAATAGGAACACCTTTGAGGAGCGTCCAAATGCTCTCCGAGAGGT... 497
165 GlyLysGluGlnThrAlaLysAsnAlaIleTriPheLysPcysGlyIleH 181
   : |||:||||| : |||:||||| : |||:|||||
498 GGAAGCGCTGCTGTAATTAATTAAGATCTTTTATAGATGTGGCTTCA 547
181 sAlaArgGluTriPheSerProAlaPheCysLeuTriPheIle..... 195
548 TGCAGAGAAATGATCTACCTGCTTTTGGCAGTGTGTGACAGAGG 597
195 ..... 195
598 CTGTCCACCTTATGAAAGAAACATTTATGACTGACTTGTGAATTAAG 647
195 ..... 195
648 TTGACTTCTATATCTGCTGCTTGAATATGACGGTTATGCTTATTC 697
196 ... GlyHisAsnArgMetLysPheGlnValSerPheTyrAlaAsnA 211
   : |||:||||| : |||:||||| : |||:|||||
698 TTGAAACAGAGTCGCATGTGAGAGAGACGCTCTGTAAATGCTGTA 747
211 snHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisLTP 227
   : |||:||||| : |||:||||| : |||:|||||
748 GTACATTCATGTTGATGATCCCAACAGAAATTT...GATGCTGATGG 794
228 CysGluGluGlnLysSerSerSerSerSerSerGlyLysTyrCysGlyLe 244
   : |||:||||| : |||:||||| : |||:|||||
795 TGCTCTGTGCTGCTTCAAGAAATCCCTGATGAGAACTATGTGCTGAG 844
244 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgA 261
   : |||:||||| : |||:||||| : |||:|||||
845 TAAACCTGAATCAGAAAGAGACCAAACTTTGGCTGACTTCATTCGAA 894
261 rGAsnIleAsnGlnIleLysAlaTyrLysSerMetHisSerLysSerGln 277
   : |||:||||| : |||:||||| : |||:|||||
895 GAAACCGCTCATATTCAGAGGCTTACTTGAACATTTATCTTACCTCCAG 944
278 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 294
   : |||:||||| : |||:||||| : |||:|||||
945 AGCTCCTTATTCATACCTTATCTTATGATCTTAACATCAACAAATA 994
294 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysHis 311
   : |||:||||| : |||:||||| : |||:|||||
995 AAACGATTTCTATGCTAAGAAAGCCATCCGAGAACTA...AAGTAC 1041
311 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
   : |||:||||| : |||:||||| : |||:|||||
1042 TGTGGCAGCAGCAATATAGTATGTCGAGAGCTGCAACAAATTTATCT 1091
328 AlaProGlyGlyLysAspThrLysPheLysPheLysLysLysLys 344
   : |||:||||| : |||:||||| : |||:|||||
1092 GCTGTGTGTGCTCAGATGATGCTTATGACGAGGTATTAATATATGC 1141
344 rPheThr 346
   : |||:||||| : |||:||||| : |||:|||||
1142 CTTCAC 1148

```

seq_name: gb:om:SSC133775

seq_documentation_block:
LOCUS SSC133775 1251 bp mRNA linear MAR 09-JUL-1999
DEFINITION Sus scrofa mRNA for pancreatic procarboxypeptidase B.
ACCESSION AJ133775
VERSION AJ133775.1 GI:5457421
KEYWORDS carboxypeptidase B.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

[illegible]

seq_name: gb-pr:HS224866
seq_documentation_block: 1310 bp mRNA linear PRI 12-MAR-1998
LOCUS HS224866
DEFINITION Homo sapiens mRNA for pancreatic procarboxypeptidase B.
ACCESSION AJ224866
VERSION AJ224866.1 GI:2959342
KEYWORDS carboxypeptidase B.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Aviles, F.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1998) Aviles F., Bioguinmica i Biologia Molecular,
MEDLINE Universitat Autnoma de Barcelona, Cerdanyola 08193, Cerdanyola
FEATURES
source 98182241 Location/Qualifiers
1..1310
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="pancreas"
5..50
6..1259
/EC_number="3.4.17.2"
/function="zymogen of pancreatic carboxypeptidase B"
/note="inactive precursor of carboxypeptidase B"
/codon_start=1
/evidence="experimental"
/product="procarboxypeptidase B"
/protein_id="CAA12163.1"
/db_xref="GI:2960072"
/db_xref="SWISS-PROT:P15086"
/translacion="MALLLVLTVALASAHGGEHEGEKEKVRVNEDEHHIIRREL
ASTQIDFWKPDVSTQIKPHSTVDERVAKEDPTVYENLKNELQYKYLISNLRVVE
AOPDSRVATGHSYEKRYKMETIETKQVATENPALISVIGTIEGRATVILKVG
KAGQNPALFMDGCFHAREWISPAFCMFREAVRTGRETQVELLNKLPDYLPVL
NIDGILTYTWTSRFRKRTSTHTGSSCIGDPNRFDMGCEIASRNPCEDEYCGPA
AESKETKALADPIRNKLSIKAVLTISYSDMMIIPYSYAVKIGENNAEELALAKAT
VKELASLHGTYTGPGATITYPAAAGSDMAVDGRIYSPTFELRDIGRGFLPES
QIRATCEETFLAIKYVASVLEHLY"
51..1256
/product="procarboxypeptidase B"
BASE COUNT 353 a 316 c 305 g 336 t
ORIGIN
Alignment_scores:
Quality: 641.00 Length: 382
Ratio: 2.514 Gaps: 7
Percent Similarity: 66.754 Percent Identity: 35.340
Alignment_block:
US-09-980-881-2 x HSA224866 ..
Align seg 1/1 to: HSA224866 from: 1 to: 1310
3 LeuCySerLeuAlaValLeuValProIleValIleuPheCyGluGlnI 19
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6 AAGTGGCAGCTCTGCTGTGACTGTGGCGCTGGCATCTGCATCA 55
19 sValIleuAla...PheGlnSerGlyGlnValIleuAlaIleuProArgT 35
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56 TGGTGGTGAGCACTTGAAGCGAGAGAGGTTCGTGTTAAGCTGAAG 105

35 hrSerArgGlnValGlnValIleuGlnAsnIleuThrThrTyrGluIle 51
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
106 ATGAATAACATTAACATTAATCCGCGAGTGGCAGACAGCCAAATP 155
52 ValIleuTrpGlnProValIleuAlaAspIleuValIleuValIleuVal 68
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
156 GACTTGTGAAGCCAGATTTGTGCACAAATCAAACTCAACACTACAGT 205
68 IHisPhePheValAsnIleSerAspValAspAsnValIleValIleu 85
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
206 TGACTTCGTTTAAAGCAAGATACCTGCACCTGTGAGCAAGTGTCTAA 255
85 snValSerGlyIleProCySerValIleuValIleuValIleuValIleu 101
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
256 AGCAGATGAACATCAATCAAGTACTGATTAAGCAACCTGAGAAATGTG 305
102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgIleSerAla 118
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
306 GTGGAGGCTCAGTTGTGATACCGGGT...CGTCACACAGACA 346
118 rTyrTyrGluGlnIleThrIleSerIleuAsnGlnIleTyrSerTrpIle 135
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
347 CAGTTATGAGAAAGTCAACAAGTGGAAACAGATAGAGCTTGACATCAC 396
135 heIleThrGluArgHisProAspMetLeuThrIleHisIleGlySer 151
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
397 AAGTCCGCTGAGAAATCCAGCCCTCATCTCTGCACTGTATTCGAGAC 446
152 SerPheGlnIleTyrProLeuTyrValIleuValIleuValIleuVal 168
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
447 ACATTGAGGAGACGCGCTATTACCTCTGAAAGTT...GGCAAACTGTG 493
168 nTrAlaIleAsnAlaIleThrIleAspGlyIleHisIleIleAlaArgGln 185
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
494 ACAAAATAAACCTGCCATTTTCATGACAGTGGTTCATGACCAAGAGT 543
185 rPleSerProAlaPheCysLeuThrPheIle... 195
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
544 GGAATTCTCTCGATCTCTGCCAGTGGTTGTAGAGAGCGTTCCTTACC 593
195 195
594 TATGACGTGAGATCCAAAGTACAGACTTCTCCAAAGTTAGACTTTTA 643
196 196
644 TGCCCGCCTGCTCTCAATATGATGATGCTACATCTACCTGGACCAAGA 693
198 snArgMetIleTrpArgIleAsnArgSerPheTyrAlaAsnHisCysIle 214
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
694 GCGGATTTTGAAGAAAGACTCGCTCCACCACTACCTGATCTAGCTGAT 743
215 GlyThrAspLeuAsnSerAsnPheValSerIleTyrHisTrpCysGluGln 231
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
744 GGCACAGACCCCAACAGAAATTT...GATGCTGTGTGTGAATTTGG 790
231 yAlaSerSerSerSerCysSerGlnThrTyrCysGlyLeuTyrProGln 248
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
791 AGCCTCTGAAACCCCTGTGATGAATCACTACTGTGACCTGCCGACAGT 840
248 erGluProGluValIleValIleValIleSerPheLeuValArgAsnIleAsn 264
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
841 CTGAAAGAGACCAAGGCCCTGGCTGATTTGATCCGCAAAACCTCTCT 890
265 GlnIleValAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 281
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
891 TCCATCAAGGCATATCTGACAAATCCACTGTACTCCCAATGATATGTA 940
281 eProTyrSerTyrThrArgSerIleSerIleValAspHisGlnIleuSer 298
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
941 CCCTTACTCATATGCTTCAAACTCGGTGAGAAACAATGCTGAGTTGAATG 990
298 euValIleSerGlnValIleValIleValIleAspIleThrSerIleAsnThr 314

```

991 CCTGGCTAAGCTACTGTGAAGAACTT...GCCTACGTGACAGCGAC 1037
315 ArgTyrThrIstgIyHstgIySerGluThrLeuTyrLeuAlaProGlyI 331
1038 AAGTACACATATGCGCCCGGAGCTACACAAATCATCTCTGCTGGGG 1087
331 yGlyAspAspTyrPileTyrAspLeuGlyIleYstYrSerPheThr 346
1088 CTCGTGACGACTGGGCTTATGACCAAGATCATATTCCTTCACC 1133

```

seq_name: gb.pr:BC015338

seq_documentation_block:

```

LOCUS      BC015338          1454 bp    mRNA    linear    PRI 04-OCT-2001
DEFINITION Homo sapiens, clone MGC:21282 IMAGE:4422620, mRNA, complete cds.
ACCESSION  BC015338
VERSION    BC015338.1  GI:15929838
KEYWORDS   MGC.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE  1 (bases 1 to 1454)
AUTHORS   Strausberg, R.
TITLE      Direct Submission
JOURNAL    Submitted (01-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA

```

```

REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: ATCC

```

CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mdickpaxill.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAX Plate: 27 Row: K Column: 21
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4503002.

FEATURES

```

source
  1..1454
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="MGC:21282 IMAGE:4422620"
    /tissue_type="Adrenal cortex, carcinoma"
    /clone_11b="NIH_MGC_84"
    /lab_host="DH10B"
    /note="Vector: PCMV-SPORT6"
    5..1258
      /codon_start=1
      /product="Unknown (protein for MGC:21282)"
      /protein_id="AAH1538.1"
      /db_xref="GI:15929839"
      /translation="MLALLVTVVALASAHGGEHEGEKVEFVNVNEDNHNITIEL
      ASTQIDEMKPSYTOIKPHSTVDFRVKAEDEVYENYVKOMELQYKVLISLRNVE
      AOFDSRVATGHSYERKYNKWTETIEMTQVATENPALTISRTIGTFEBERAYILKVG
      KAGQNKPAIFMPCGFHARBMIPARCOMFVRARVTRTGREIVYTEILDLDYFVPLV
      NIDGTYTWTKSRFWKTRSTHTGSSCGTDPNRNPDACWCEIGASRNPCDFTYCGPA
      ASEKETRALDPIRNKLSSIKAYLLTISYSQMMIYPYSYAVKLGESNNALNALMAT
      VKEIASLHGTGTYGATTIYPAAAGSDMAVDGIRYSFTFELDRTGRYGFILPES
      QIRATCEEFELAIKRYVAVYLEHLV"

```

CDS

```

BASE COUNT  398 a      335 c      327 g      394 t
ORIGIN

```

```

alignment_scores:
  Quality: 641.00      Length: 382
  Ratio: 2.514        Gaps: 7
  Percent Similarity: 66.754      Percent Identity: 35.340

```

alignment_block:

US-09-980-881-2 x BC015338 ..
 Align seg 1/1 to: BC015338 from: 1 to: 1454

```

3 LeucylSerLeuAlaValLeuValProIleValLeuPheCysGluGlnI 19
  ::::::::::::::::::::|
5 ATGTGGACACTCTGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTATCA 54
  ::::::::::::::::::::|
19 sValPheAla...PheGlnSerGlyGlnValLeuAlaLeuProArgT 35
  |::::::::::::::::::|
55 TGGTGTGACGACCTTGTGAAGCGAGAGAGTGTCTCCGTTCAGCTTGAG 104
  ::::::::::::::::::::|
35 hrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyrGluIle 51
  ::::::::::::::::::::|
105 ATGAATAATCACATTATACATATCCGCGAGTGGCCAGCAGCAGCAGAT 154
  ::::::::::::::::::::|
52 ValLeuTyrGlnProValThrAlaAspLeuIleValIstYstYsGlnVa 68
  ::::::::::::::::::::|
155 GACTTCTGGAAGCCAGATTCTGTCTCACCACAAATCAACCTCAGTACAGT 204
  ::::::::::::::::::::|
68 IHisPheValAsnAlaSerAspValAspAsnValIstYstYsLeu 85
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
205 TGACTTCCTGTAAAGCAGAGATCTGTCTGCTGAGAGATGTTCTAA 254
  ::::::::::::::::::::|
85 snValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
  ::::::::::::::::::::|
255 AGCAGAAATGACATACATACAGAGTACTGATACCAACCTGAGAAATGTG 304
  ::::::::::::::::::::|
102 IlleGlnGlnIleSerAsnAspThrValSerProAlaIleAlaIle 118
  ::::::::::::::::::::|
305 GTGAGAGCTCAGTTTGATGAGCGCGGT.....CTGCACAGCAGACA 345
  ::::::::::::::::::::|
118 rTyrTyrGluGlnIleThrIstYstYsLeuAlaGluIleTyrSerIleGlu 135
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
346 CAGTTATGAGAGATACACAGAGTGGAGAGCGATGAGCTGCTGACTAAC 395
  ::::::::::::::::::::|
135 IleIleThrGluArgHisProAspMetLeuThrIstYsIleHisIleGlySer 151
  ::::::::::::::::::::|
396 AAGTCGCACGAGATACACAGCTCTGCTGCGAGGTATATGGAAAC 445
  ::::::::::::::::::::|
152 SerPheGluIstYrProLeuTyrValLeuIstYsValSerGlyIstYsGlu 168
  ::::::::::::::::::::|
446 ACATTTGAGGAGCGCGCTATTACCTCTGCTGAGAGT...GCCAAAGCTGG 492
  ::::::::::::::::::::|
168 nThrAlaIstYsAsnAlaIleThrPheIstYsGlyIleHisIstYsGlu 185
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
493 ACAAAATTAAGCTCCATTTTCATGAGCTGTGTTCCATGCCAGAGAGT 542
  ::::::::::::::::::::|
185 rPheSerProAlaPheCysLeuTyrPheIle..... 195
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
543 GGATTTCTCTGCAATTCGCCAGTGGTTTGTAAAGAGAGCTGTCTGAC 592
  ::::::::::::::::::::|
195 ..... 195
593 TATGAGCTGAGATCAAGTACAGAGCTTCTGCACAGTTAGACTTTTA 642
  ::::::::::::::::::::|
196 .....GLYHis.....A 198
  ::::::::::::::::::::|
643 TTGCTCTGCTCTGCTCAATATGATGCTCATATCTACACCTGACCAAGA 692
  ::::::::::::::::::::|
198 snArgMetTyrArgIstYsAsnArgSerPheTyrAlaAsnAsnHisCysIle 214
  ::::::::::::::::::::|
693 GCCGATTTTGGAGAAAGACCTGCTGACCATATGATGATGATGATGAT 742
  ::::::::::::::::::::|
215 GlyThrAspLeuAsnSerAsnPheValSerIstYsIstYsGluGlu 231
  ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

50 rovaltlnfalaaspleullevallysllyslcInvalhlsphpheval 72

128	CAGATTCTGTGCACAAATCAAACTCAACCTACAGTACAGTTGACTTCCTGGTGC	177
73	AsnAlaSerAspValAspAsnValAlaHisLeuAsnValSerGlyI	89
178	AAACCAAGAAATATTTTGGCTGTGGAGACACTTCTGGACAGAAATGAAGT	227
89	eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI	106
228	ACAAATATGAGGTACTCATAAACAACCTGAGATCTGTCTCTGGAGGCTAGT	277
106	LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln	122
278	TTTGACACAGAGCTC.....CGTCAACTGGACACAGATTATGAGACG	318
123	TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr	139
319	TCACACACACTGGGAACAGATCGAGGCTTGGACTTGAAGCAAGTACCAAGTGA	368
139	GHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluYst	156
369	AAATCCAGACCTCATCTCTGCGACAGGCACTCGGAACATCTTTTGGAA	418
156	YrProLeuTyrValLeuLysValSerGlyGluGlnThrAlaLysAsn	172
419	ACAATATATCTCTCTCAAGTT.....GCGAAACCTGGACCAAAATGAAGCCT	465
173	AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl	189
466	GCCATTTCATGAGCACTGTGGTTCCATGCCAGAAATGATTTCCCATGC	515
189	AspPheCysLeuTrpPheLe.....	195
516	ATTTTCGCAAGTGGTTGTGTAGAGAGGCTGTCTCAACCTATGATATGAGA	565
195	195
566	GTCACATGACAGAAATTCCTCAACAACCTAGACTTTATGTCTTGGCTGGC	615
196GlyHis.....AsnArgMetIleTrpAr	202
616	CTCAATATATGATGGCTACATCTACACCTGGACCAAGCAAGATGGAG	665
202	GlyAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyTrpAspLeuA	219
666	AAAGACCCCTGTACCAACAGCTGGAACCTGCTCATGGCAACAGACCCCA	715
219	snSerAsnPheValSerLysHisTrpCysGluGlnGluValaSerSerSer	235
716	ACAGAAATTTT...GATGCTGGGTGGTCGACAACTGGAGCCTCTACAGAC	762
236	SerCysSerGlnThrTyrTrpGlyLeuTyrProGlnSerGlnProGluVa	252
763	CCCTGGAGTAGAGACTTACGTGGATCTGTCTGCAAGACTGTGAAAAAGAGAC	812
252	LysAlaValAlaSerPheLeuArgTrpAsnIleAsnGlnIleLysAlaTr	269
813	CAAGGCGCTGGGTGATTTATACGCAACAACCTCTCTCATCAAAAGCAT	862
269	YrIleSerMetHisSerTyrSerGlnHisIleValaPheProTyrSerTyr	285
863	ACCTGACAGTCCACATCTACTCAGAGATATCTACTACCCCTATATCTAT	912
286	ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG	302
913	GATTCACAACTCCCGAGAACAAATGCTGAGTTGAATTAACCTGGCTAAAGC	962
302	UAlaValAlaGalaIleAspLysThrSerLysAsnThrArgTyrTrpHisG	319
963	TTCCTGTGAAGAAGACTT...GCTACACTGTATGCGACCAAGTACATACAG	1009
319	LysHisGlySerLysIleuThrLeuTyrLeuAlaProGlyGlyLysAspAr	335
1010	GCCCAAGAGACTCAACAATATCTCTGCTGCTGGGGGCTGTGTATACACGG	1059

336 IleTyrAspLeuGlyIleLysTyrSerPheThr 346
|||||
1060 GCTTATGACCAAGGAAATCAATATTCCTCACC 1092

This Page Blank (uspto)

OM of: US-09-980-881-2 to: N_Geneseq_032802:* out_format : pfs

Date: Sep 18, 2002 7:02 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-O=/cgn2_1/USPTO_pool/US09980881/runat_16092002_140130_10714/app-query.fasta.1.1126
-DB=N_Geneseq_032802 -QFMT=fastp -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPT=0.000 -LOOEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELAP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=ptc
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ps
-NORTEXT=HEAPSITE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 -CGN1_1.0 -NCPV=6 -ICPV=3 -LONGLOG
-DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-2
Query length: 360
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 386.310000

```

score_list:
Sequence      Strid Orig      ZScore      EScore Len 1 Documentation
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC81962 + 1911.00 3685.66 6.0e-197 15
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV74302 + 1832.00 3534.43 1.6e-188 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV74301 + 1832.00 3530.75 2.6e-188 17
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAV74300 + 1832.00 3530.75 2.6e-188 17
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74303 + 1832.00 3530.75 2.6e-188 17
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAV74304 + 1827.00 3521.03 8.9e-188 17
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV74305 + 889.00 1698.91 2.8e-86 14
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAV74306 + 640.50 1217.15 1.9e-59 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV74307 + 639.00 1213.78 2.9e-59 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV74308 + 639.00 1213.78 2.9e-59 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV74309 + 631.50 1199.00 1.9e-58 12
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74310 + 623.00 1176.49 3.5e-57 21
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74311 + 611.50 1159.86 2.9e-56 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV74312 + 604.00 1149.27 1.1e-55 92
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV74313 + 576.00 1094.88 1.2e-55 92
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74314 + 564.00 1061.85 4.7e-51 4
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74315 + 563.00 1067.28 4.2e-51 4
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74316 + 563.00 1067.28 4.2e-51 4
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74317 + 563.00 1065.24 5.5e-51 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74318 + 563.00 1065.24 5.5e-51 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74319 + 563.00 1061.68 8.7e-51 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV74320 + 561.00 1064.76 5.8e-51 99
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV74321 + 561.00 1064.15 6.3e-51 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV74322 + 554.00 1050.47 3.7e-50 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74323 + 554.00 1050.47 3.7e-50 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV74324 + 554.00 1050.47 3.7e-50 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74325 + 553.00 1048.52 4.7e-50 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74326 + 553.00 1048.52 4.7e-50 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74327 + 550.00 1042.69 9.9e-50 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74328 + 548.00 1038.80 1.6e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74329 + 548.00 1038.80 1.6e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74330 + 547.00 1036.85 2.1e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74331 + 546.00 1034.91 2.7e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74332 + 546.00 1033.46 3.2e-49 1
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74333 + 545.00 1032.96 3.4e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74334 + 545.00 1032.96 3.4e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74335 + 545.00 1032.96 3.4e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74336 + 543.00 1029.07 5.7e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74337 + 543.00 1029.07 5.7e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74338 + 543.00 1029.07 5.7e-49 10
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74339 + 543.00 1022.50 1.3e-48 18
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV74340 + 543.00 1022.50 1.3e-48 18

```

```

/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74341 + 542.00 1027.13 7.3e-49
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV74342 + 540.00 1023.23 1.2e-48
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74343 + 528.50 1000.96 2.1e-47
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAV74344 + 519.50 981.35 2.6e-46
/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAV74345 + 519.50 980.89 2.7e-46

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAV74346

seq_documentation_block:

```

ID AAC81962 standard; CDNA; 1573 BP.
XX
AAC81962:
XX
01-MAR-2001 (first entry)
XX
Human brain carboxypeptidase B cDNA.
XX
Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
XX cerebroprotective; antialzheimers; neurotrophic; neuroprotective;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma; ss.
XX
Homo sapiens.
XX
Key Location/Qualifiers
XX CDS 18..1100
XX FT /*tag= a
XX FT /product= "carboxypeptidase B"
XX
XX WC20006617-1.1.
XX
09-NOV-2000.
XX
01-MAY-2000; 2000MO-JP02878.
XX
30-APR-1999; 99UP-0125169.
XX
(MATS/) MATSUMOTO A.
XX
Matsumoto A:
XX WPI; 2000-687534/67.
XX P-PSDB; AAB11457.
XX
Human brain carboxypeptidase B isolated from the hippocampus useful for
XX screening agents for the treatment of Alzheimer's and other brain
XX disorders -
XX
Claim 2b: Page 64-68; 84pp; Japanese.
XX
This invention describes a novel protein with peptidase activity
XX against brain beta-amyloid precursor protein which has been isolated from
XX human hippocampus and which has cerebroprotective, antialzheimers,
XX neurotrophic, neuroprotective and hemostatic activity and which can be used
XX as a vaccine or for gene therapy. The protein, and compounds identified
XX by screening as promoters or inhibitors of its activity, are used to
XX regulate beta-amyloid accumulation in the brain and treat or prevent
XX diseases in which this occurs, such as Alzheimer's, senile dementia,
XX inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;
XX

```

alignment_scores:

Quality: 1911.00 Length: 360
Ratio: 5.308 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-980-881-2 x AAC81962 ..
Align seg 1/1 to: AAC81962 from: 1 to: 1573

```

1 MetLysLeuGysSerLeuAlaValLeuValProIleValLeuPheCysGly 17
  |||
18 ATGAAGCTTTCAGCCCTTCAGCTCTTGTACCCATTGCTCTCTCTGTA 67
  |||
21 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaIleProA 34
  |||
68 GCAGATGCTCTCGGTTCCAGAGGCGCAAGTCTAGCTGCTCTCTTA 117
  |||
34 rGthSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrGlu 50
  |||
118 GAACCTTCAGGCAAGTTCAGTCTACAGAACTTACTACAACTATGAG 167
  |||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGly 67
  |||
168 ATTGTTCTCTGCGACCGCGTAACAGCTGACTTATGTAAGAAAAACA 217
  |||
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
  |||
218 AGTCCATTTTGTGTAATGATCATGTGATGCAATGTGAAACCCATT 267
  |||
84 euAsnValSerGlyIleProCysSerValLeuAlaAspValGluAsp 100
  |||
268 TAAATGTGAGCGGAAATTCATGAGTCTGCTGCGACAGCTGGAAGAT 317
  |||
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
  |||
318 CTATTTCACACGAGATTTCCACAGACAGTCAGCCCGAGCCTCGC 367
  |||
117 aSerTyTyGluGlnTyThrHisSerLeuAsnGluIleTySerTrpIleG 134
  |||
368 ATGCTACTAGAACAGATACACTGCTAAATGAAATCTATCTTGATAG 417
  |||
134 LuPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
  |||
418 AATTATTAATGAGAGGATCTGATATGCTTACAAAAATCCACATGGA 467
  |||
151 SerSerPheGluLysTyProLeuTyValLeuLysValSerGlyLysGly 167
  |||
468 TCCATCATTTAGAAAGTACCCACTGATGTTTAAAGGTTTCTGGAAGA 517
  |||
167 uGlnIlePheAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
  |||
518 ACAAAACAGCCAAAATGCCATATGATGATGCTGGAATCCATGCCAGAG 567
  |||
184 LuTrpIleSerProAlaPheCysLeuTrpPheIleGlnHisAsnArgMet 200
  |||
568 AATGGATCTCTCTGCTTCTGCTGTTGTTGATAGGCCATTAATCGAATG 617
  |||
201 TrpArgLysAsnArgSerPheTyraAlaAsnAsnHisCysIleGlyThrAs 217
  |||
618 TGGAGAAAGAACCGTTCTTGTATGGCAACATTCATTCGGAACAGAG 667
  |||
217 PleuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerS 234
  |||
668 CCGAATATACAACTTGTCTCCAAACAGTGTGTGAGGAAGGTCATCA 717
  |||
234 SerSerSerCysSerGluTrpTyTyCysGlyLeuTyProGlnSerGluPro 250
  |||
718 GTTCCCAATGCTCGAAGAACCTACTGTGACTTATCTGATGTCAGAACCA 767
  |||
251 GluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLeu 267
  |||
768 GAAGTGAAGCAGAGTGGCTAGTTCTTGAGAGAAATATCAACACAGATTAA 817
  |||
267 bAlaTyTrIleSerMetHisSerTySerGlnHisIleValPheProTyrs 284
  |||
818 ACATATACATCAGATGATTCATACCCACAGCATATAGTGTTCATAT 867
  |||
284 eTyTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
  |||
868 CCTATATACGAGATGAAGCAAAAGACATGAGAACTGTCTCTAGTAGCC 917
  |||

```

```

301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyThr 317
  |||
918 AGTGAAGAGATTGCGTATATGACAAACCTAGTAAATATACAGATATAC 967
  |||
317 rHisGlyHisGlySerGluThrLeuTyThrLeuAlaProGlyGlyAspA 334
  |||
968 ACATGGCCATGCGTCAGAAACCTTATACCTAGCTCTGAGAGTGGGAGC 1017
  |||
334 sPTripIleTyAspLeuGlyIleLysTySerPheThrSerAsnProPro 350
  |||
1018 ATTGATCTATGATTTGGGCATCAAAATATTCGTTTACATCAAAACCCAC 1067
  |||
351 ValGluLysLeuLeuProLeuSerLeuLys 360
  |||
1068 GTAGAGAAAGCTTTTGGCGCTGCTCTTAA 1097
  |||
seq_name: /STD1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV74302
seq_documentation_block:
ID   AAV74302 standard; cDNA; 1272 BP.
XX
AC   AAV74302;
XX
DT   28-APR-1999 (first entry)
XX
DE   Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
KW   Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
XX   polymorphism detection; thrombotic disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO9855645-A1.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998; .98WO-EP03244.
XX
PR   03-JUN-1997; 97US-0869057.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Morser MJ, Nagashima M.
XX
DR   WPI: 1999-045800/04.
XX   P-PSDB: AAW92270.
XX
PT   Detecting new polymorphism of human plasma carboxypeptidase B -
PT   comprises Alanine or Threonine at position 147 of protein by DNA or
PT   protein analysis, useful to detect risk of thrombotic disease in
PT   humans
XX
PS   Example 1; Page 24; 35pp; English.
XX
CC   This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC   hPCPBthr147. The invention relates to a method for determining the
CC   presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC   comprises obtaining a prepared tissue or blood sample and determining the
CC   presence of DNA coding for naturally occurring polymorphs of the protein
CC   containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC   respectively). Determination of the relative distribution of the PCPB
CC   polymorphs in a patient's blood by genetic or protein analysis by the
CC   methods is useful to determine the risk of thrombotic disease in humans.
CC   Such assessments may be made by accumulating information concerning the
CC   relative distribution of the different polymorphs within the general
CC   population compared with populations known to be at risk and establishing
CC   a PCPB polymorph profile for at-risk patients.
XX
SO   Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

```

alignment_scores: Length: 415
Quality: 1832.00

Ratio: 5.103 Gaps: 3
Percent Similarity: 86.506 Percent Identity: 85.783

alignment block:
US-09-980-881-2 x AAV74302 ..

Align seg 1/1 to: AAV74302 from: 1 to: 1272

```

1 MetIysLeuCySerSerLeuAlaValIleuValProIleValIleuPheCysGI 17
1 ATGAGGCTTTCAGACCTTGCAGCTCTGTACCATCTGCTCTCTCTGTA 50
17 uGlnHisValPheAlaPheGlnSerGIyGlnValIleuAlaIleuProA 34
51 GCAGCAGTCTTCGCTTTCAGATGGCCAAAGTTCAGCTGCTCTCTCTTA 100
34 rGhrSerArgGlnValGlnValIleuGlnAsnLeuThrThrThyGlu 50
101 GAACCTTCAGCAGATTCAAGTCTTACAGAACTTACTACACATATAGAG 150
51 IleValIleuTrpGlnProValThrAlaAspLeuIleValIlySlySGI 67
151 ATGTCTCTGGCAGCGGTACAGCTGACCTTATGTGAAGAAAAACA 200
67 nValHisPhePheValAsnAlaSerAspValAspAsnValIlyAlaHisL 84
201 AGTCATATTTTGTAAATGATCATCTGATGTGACAAATGTGAAGCCCAT 250
84 euAsnValSerGIyIleProCySerValIleuAlaAspValGluAsp 100
251 TAAATGGAGCGGAATTCATGAGCTGTCTGTGGCAGAGCTGGAAGAT 300
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
301 CTATTCAGACAGCAGATTTCACAGCAGACAGTCCGCCGAGCTCGGC 350
117 aSerTyTyTyGluGlnIlyHisSerIleuAsnGluIleTySerTrpIleG 134
351 ATGCTACTATGAAACAGTATCATCTCACTAAATGAAATCTATCTTGATG 400
134 IupheIleThrGluArgHisProAspMetLeuThrIlyHisIleGly 150
401 AATTTAACTGAGAGCAGCATCTGATATGCTTACAAAATCCACATTGGA 450
151 SerSerPheGluIlyTyTyProLeuTyValIleuIlyValSerGIySGI 167
451 TCCGCAATTGAGAAGTACCCACTCTATGTTTAAAGTTCTCGAAAAAGA 500
167 uGlnThrAlaIlyAsnAlaIleTrpIleAspCySGIyIleHisAlaArgG 184
501 ACAAAACGCCAAAATGCAATATGATGATGACTGTGGAATCCATGCCAGAG 550
184 IuTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
551 AATGGATCTCTCCGCTTCTGCTTGCTTGCTTCAATGAGCCATATACCTCA 600
197 ..... 197
601 TTCATGAGTAATATAGGCAATATACCAATCTCTGAGGCTGTGGAATTT 650
197 ..... 197
651 CTATGTTATCCGGTGGTTAATGTGACGGTTATGACTACTCATGAGAAA 700
198 ..AsnArgMetTrpArgIlyAsnArgSerPheTyAlaAsnHisCys 213
701 AGAATCCAAATGTGAGAAAGACCGTCTTCTATGCGAACAATCATTTCC 750
214 IleGIyThrAspLeuAsnSerAsnPheValSerIlyHisIstrPySGIuGI 230
751 ATCGAACAAGACCTGAATATGAGAACTTGTCTCAAAACACTGTGTGAGGA 800
230 uGIyAlaSerSerSerSerGIyIuThrTyCySGIyLeuTyTrpG 247

```

```

|||||
801 AGTGCAATCCAGTTCTTCATGCTCGAAACCTACTGTGAGACTTATCCTG 850
247 IuSerGIyProGIuValIlyAlaValAlaSerPheLeuArgArgAsnIle 263
|||||
851 AGTCAGAACCCAGAGTGAAGCAGTGGCTGACTTCTTTCAGAGAAATATTC 900
264 AsnGlnIleIlyAlaTyTrIleSerMetHisSerTySerGIuHisIleVa 280
901 AACCAAGATTAAAGCATCATCAGCATGATCATATCTCCAGCATATAGT 950
280 IpheProTySerTyThrArgSerIlySerIlyAspHisGIuGluLeuS 297
951 GTTTCATATTCATATACAGAAAGTAAAGCAAGACCATGAGAACCTGT 1000
297 eIleuValAlaSerGIuAlaValArgAlaIleAspIlyThrSerIlyAsn 313
1001 CTCTAGTAGCCAGTAGAACAGTTCGTGCTATTTGAGAAACTAGTAAAT 1050
314 ThrArgTyThrHisGIyHisGIySerGIuThrLeuTyIleuAlaProGI 330
1051 ACCAGGTATACATGAGCAGTGGCTCAGAAACCTTATACCTAGCTCTGG 1100
330 yGIyGIyAspAspTrpIleTyAspLeuGIyIleIlyTySer Phe... 345
1101 AGTGGGAGCAGATTGATCTATGATTTGGCCATCAAAATATTCGTTACAA 1150
346 .....Th 346
1151 TTGAACTTCAGAAATACGGGCAATACGAGATTCTTGCTCGCGAGCGTTAC 1200
346 rSerAsnProProValGIuIlySlyLeuLeuProIleuSerIleuIlyS 360
1201 ATCAAAACCACCTGTAGAGAGCTTTGGCGCTGTCTCTATAAA 1243
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NAI196.DAT:AA11671
seq_documentation_block:
ID AAT11671 standard; DNA: 1749 BP.
NC AAT11671;
XX
DT 12-APR-1996 (first entry)
XX
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 41..1312
FT /tag= a
FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT /tag= b
FT mat_peptide 107..1309
FT /tag= c
XX
PN US5474901-A.
XX
PD 12-DEC-1995.
XX
PF 01-FEB-1991: 91US-0649591.
XX
PR 01-FEB-1991: 91US-0649591.
PR 14-OCT-1992: 92US-0959944.
PR 15-DEC-1993: 93US-0167727.
PR 19-JUL-1994: 94US-0277540.
XX
PA (GETH ) GENENTECH INC.
XX

```

PI Drayna DT, Eaton DL;
 XX WPI: 1996-039508/04.
 DR P-PSDB; AAR90293.
 XX
 XX
 PT Antibody to human plasma carboxypeptidase B - useful for detecting
 PT haemophilia A
 PS
 XX Disclosure; Figure 4; 40pp; English.
 XX
 CC An antibody which specifically binds human plasma carboxypeptidase B
 CC (hPCPB) and does not cross react with other carboxypeptidases is
 CC useful for the detection of hPCPB in vitro. The antibody is also
 CC used for purifying hPCPB from a sample. Purification comprises
 CC passing a sample thought to contain hPCPB over either a column to
 CC which antibody has been bound, or a plasminogen affinity column,
 CC eluting the column and then recovering the fraction containing the
 CC hPCPB.
 XX
 SQ Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

alignment_scores:
 Quality: 1832.00 Length: 415
 Ratio: 5.103 Gaps: 3
 Percent Similarity: 86.506 Percent Identity: 85.783

alignment_block:
 US-09-980-881-2 x AAT11671 ..

Align seg 1/1 to: AAT11671 from: 1 to: 1749

```

1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
41 ATGACCTTTGGCAGCCTTGACGTCCTTGACCATGTTCTCTGCTGA 90
17 uGlnHisValPheAlaPheGlnSerGlyGlnValIleAlaIleProA 34
91 GCAGCATGCTTGGCGTTTCAGAGTGCCAAAGTTCAGCTCTTCCTA 140
34 rGthSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyGlu 50
141 GAACCTCTAGGCAAGTTCAGATCTACAGATTTACTACACATATGAG 190
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
191 ATGTTCTCTGGCAGCGGTAACAGCTGACTTATTTGTAAGAAAAACA 240
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
241 AGTCCATTTTGTAAATGCAATGATGTCGCAATGTGAAGCCCAT 290
84 euAsnValSerGlyIleProCySerValIleuAlaAspValGluAsp 100
291 TTAATGTGAGCGAATTCATGCAAGTCTGCTGGCAGACGGAAGAT 340
101 leuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
341 GTTATTCACACAGCATTTTCCAAACGACACATCAGCCCGGAGCTCC 390
117 aSerTyTyGlnGlnIleThrHisSerLeuAsnGluIleTySerTrpL 134
391 ATCGTACTATGAAAGTATCACTCACTAAATGAAATCTATTTTGGATG 440
134 IuPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
441 AATTATTAAGTGAAGAGCATCTGATATGCTTAACAAAAATCACAFTG 490
151 SerSerPheGlnLysTyProLeuTyValLeuLysValSerGlyLysG1 167
491 TCCTCATTTGAGAGTACCACTCTATGTTTAAAGTTTTCGAAAAAGA 540

```

```

167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
541 ACAAAACAGCCAAAATGCAATGATGATGATGATGATGATGATGATG 590
184 IuTrpIleSerProAlaPheCysLeuTrpPheIleGlnHis..... 197
591 AATGATCTCTCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 640
197 ..... 197
641 TTTGATGGATTAATAGGCAATATACCAATCTCTGAGCTTGGATTT 690
197 ..... 197
691 CTATGTTATCCCGTGGTTAATGAGCGTTTATGACTACTATGAAAA 740
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnAsnHisCys 213
741 AGAATCGAATGTGAGAAAGAACCGTTCCTTATGCGAAACATCATTC 790
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGlu 230
791 ATCGAACAAGCCTGAATAGAACTTGGCTTCCAAACACTGGTGTGAGA 840
230 uGlyAlaSerSerSerSerCysSerGluThrTyCysGlyLeuTyProG 247
841 AGGTGATCCAGTTCCTCATGCTCGGAAACCTATGAGACTTTATCCG 890
247 IuSerLupProGluValLysAlaValAlaSerPheLeuArgArgAsnI 263
891 AGTCAGAACCAAGTGAAGGAGCGTGTGCTTCTTGAAGAAATATTC 940
264 AsnGlnIleLysAlaTrpIleSerMetHisSerTySerGlnHisIle 280
941 AACCAATTAAGCAATACATCAGATGCAATTCATCTCCACCATATAT 990
280 IuPheProTySerTyThrArgSerLysSerLysAspHisGluGluLeu 297
991 GTTTCATATTCCTTATACAGAAATGAAGCAAGCAAGCAAGCAAGT 1040
297 erLeuValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerL 313
1041 CTCTGATACCGAGTGAAGCGATTCGCTATTTGAAACAACTATGTA 1090
314 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyLeuAlaProG 330
1091 ACCAGGTATACATGAGCCATGGCTCAGAAACCTTATACCTAGCTCC 1140
330 yGlyGlyAspAspTrpIleTyAspLeuGlyLysTySerPhe... 345
1141 AGGTGGGAGCATTTGATCTATGATTTGGCATCAAAATATTCGTTACA 1190
346 .....Th 346
1191 TTGAACCTCGAATACGGGCAATACGCAATCTTCTCCGAGCGTTAC 1240
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
1241 ATCAAAACCACTGTAGAGAGAGCTTTGCCGCTGCTCTTAAA 1283
seq_name: /SIDS1/gcdata/geneseq/geneseq-emb1/NA1997.DAT:AA62846
seq_documentation_block:
ID AAT62846 standard; DNA: 1749 BP.
XX
AC AAT62846;
XX
DT 08-NAV-1997 (first entry)
XX
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW plasma; plasminogen; ss.

```

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 41..1312
XX FT /tag= a
XX FT /product= Human PCPB
XX FT sig_peptide 41..106
XX FT /tag= b
XX FT mat_peptide 107..1309
XX FT /tag= c
XX FT misc_binding 134..177
XX FT /tag= d
XX FT /bound_molecly= 46_bp_probe
XX PN US5593674-A.
XX PD 14-JAN-1997.
XX PF 01-FEB-1991; 91US-0649591.
XX PR 01-FEB-1991; 91US-0649591.
XX PR 14-OCT-1992; 92US-0959944.
XX PR 15-DEC-1993; 93US-0167727.
XX PR 19-JUL-1994; 94US-0277540.
XX PR 27-APR-1995; 95US-0430787.
XX PA (GENTECH ) GENENTECH INC.
XX PI Drayna DF, Eaton DL;
XX DR WPI; 1997-099413/09.
XX DR P-PSDB; AAM14733.
XX XX
XX PT Using human plasma carboxypeptidase B in blood coagulation - is
XX PT functionally related to carboxypeptidase A and pancreas
XX PT carboxypeptidase B
XX PS Example 2; Column 37-42; 39pp; English.
XX CC This sequence encodes human plasma carboxypeptidase B (PCPB) which
XX CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
XX CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
XX CC purified from human plasma or by transformed cell culture by
XX CC extraction using plasminogen bound to a solid phase.
XX SO Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

alignment_scores:
Quality: 1832.00 Length: 415
Ratio: 5.103 Gaps: 3
Percent Similarity: 86.506 Percent Identity: 85.783

alignment_block:
US-09-980-881-2 x AAT62846 ..
Align seg 1/1 to: AAT62846 from: 1 to: 1749

1 MetLysLeuCySserLeuAlaValLeuValProIleValIleuPheCysG1 17
|||||
41 ATGAGGCTTGGACCTTGGACGCTCTGTACCCATGTCTCTCTGTGGA 90
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaIleuProA 34
|||||
91 GCACAGTCCTCTCCGCTTTCAGAGGCCCAAGTCTACGCTCTCTCTA 140
|||||
34 rGthrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThyGlu 50
|||||
141 GAACCTCTAGGCAAGTCAAGTCTACAGAACTTACTACAAACATATAG 190
|||||
51 IleValIleuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
|||||

```

```

191 ATTGTTCTCTGGACCGGTAACAGCTGACCTTATTGTGAAAGAAAAACA 240
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
|||||
241 AGTCATTTTCTTGTAAATGCATCTGATGTGCACAAATGTGAAGCCCAT 290
|||||
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
|||||
291 TAAATGTAGCGGAATTCATCTACATGCTCTCTGCGACAGCTGGAGAT 340
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaIleSer 117
|||||
341 CTATTTCAACACAGATTTCCACAGACAGATCCCGCCAGCCCTCCGC 390
|||||
117 aSerTyTrpGlnGlnIleTyHisSerLeuAsnGluIleTySerTrpIleG 134
|||||
391 ATCGTACTATGACAGATATCACTCACTAAATATATCTATCTTGGAATG 440
|||||
134 IupheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||
441 AATTATTAACGTAGAGCATCTGTATATGCTTACAAAATCCACATTGGA 490
|||||
151 SerSerPheGluLysTyProLeuTyValLeuLysValSerGlyLysG1 167
|||||
491 TCCCTATTTGAGAAAGTACCCTCTATGTTTAAAGTTCTGGAAGAAGA 540
|||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
|||||
541 ACAAAACAGCAAAAATGCCATATGATGATGACGTGGATCCATGCCAGAG 590
|||||
184 IuTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
|||||
591 AATGATCTCTCCCTCTTCTCTGTGTCATGAGCATATTAATCA 640
|||||
197 ..... 197
641 TTTATGGAATATAGGCGAATATACCAATCTCTGAGCGTTGTGATTT 690
|||||
197 ..... 197
691 CTATGTTATGCCGGGTGTTAATGTGAGCGTTATGACTACTATGAAAA 740
|||||
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnHisCys 213
|||||
741 AGAATCGAATGTGAGAGAAAGAACCGTCTTCTATGCGAACAATCATTC 790
|||||
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGlu 230
|||||
791 ATCGGAACAGACCTGAATAGCAACTTTCCTCCAAACACTGGTGTAGGA 840
|||||
230 uGlyAlaSerSerSerCysSerGluThrTyrcysGlyLeuTyProG 247
|||||
841 AGTGCAATCCAGTTCTCATGCTCGGAAACCTACGTGTGACTTTATCTG 890
|||||
247 IuSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
|||||
891 AGTCAGAACCAAGTGAAGCGAGCTAGCTTCTTGAGAGAAATATTC 940
|||||
264 AsnGlnIleLysAlaTyrlleSerMetHisSerTySerGlnHisIleVa 280
|||||
941 AACCGATTAAAGCATATACATGACATTCATATCTCCACACAAATATAG 990
|||||
280 IPherProTySerTyTrpThrArgSerLysSerLysAspHisGluLeuS 297
|||||
991 GTTTCATATTCCTTATACAGAAAGTAAAGCAAAACCATGAGAACTGT 1040
|||||
297 eLLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsn 313
|||||
1041 CTCTAGTAGCCAGTGAACAGAGTCTGCTATATGAGAAAATCTGTAATA 1090
|||||
314 ThrArgTyTrpHisGlyHisGlySerGluThrLeuTyLysAlaIlePro 330
|||||
1091 ACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGG 1140
|||||

```

```

330 YG|YGLVAPASPTPIET|YRASPLeuGlyLLeLysTyrSer... 345
|||||
1141 AGGTGGGACGATTTGATCTATGATTTGGCATCAATATTCGTTTACAA 1190
|||||
346 .....Th 346
1191 TTGAACCTTCGAGATACGGGACATACGATTCCTGCTCCGAGCGTTAC 1240
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
|||||
1241 ATCAAAACCCACCTGTAGAGAGCTTTCCGCTGCTCTAATAA 1283
seq_name: /STDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:AAQ41001
seq_documentation_block:
ID AAQ41001 standard: cDNA; 1749 BP.
XX
AC AAQ41001;
XX
DT 24-AUG-1993 (first entry)
XX
DE Human plasma carboxypeptidase B gene.
XX
KM PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT sig_peptide 41..106
FT mat_peptide 107..1312
FT /tag= a
FT /tag= b
FT /product= PCPB
FT /tag= C
FT /function= probe
FT /note= "used to obtain full-length clones"
XX
PN US5206161-A.
XX
PD 27-APR-1993.
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
XX
PA (GETH ) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
DR WPI: 1993-151724/18.
DR P-PSDB; AAR36273.
XX
PT New human plasma carboxypeptidase B - used as haemostatic
PT regulator for clotting blood, partic. for treating blood clotting
PT disorders, e.g. haemophilia
XX
PS Disclosure: fig 4; 40pp: English.
XX
CC Human plasma carboxypeptidase B was isolated from human plasma and
CC partially sequenced. Oligonucleotide primers were designed based on
CC the partial amino acid sequences. The primers were used in a PCR
CC amplification to identify cDNA encoding PCPB from a human liver cDNA
CC library. The PCR product was capable of encoding the first 37 amino
CC acids of PCPB; a 46mer probe was used to obtain the full-length
CC sequence which, although disclosed in the specification, is not
CC claimed. PCPB inhibits the enzymatic conversion by tPA of
CC plasminogen to plasmin in the presence of fibrinogen.
XX
SO Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

```

```

alignment_scores:
  Quality: 1827.00      Length: 415
  Ratio: 5.103
  Percent Similarity: 86.265      Percent Identity: 85.542
alignment_block:
US-09-980-881-2 x AAQ41001
Align seg 1/1 to: AAQ41001 from: 1 to: 1749
1 MetLysLeuCySserLeuAlaValLeuValProLLeuValLeuPheCySg1 17
|||||
41 ATGAAGCTTTCACACCTTCGACGTCCTGTACCCCATTTCTCTCTCTGTGA 90
17 UGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuPro 34
|||||
91 GCAGCATGCTCTCCGCTTTCAGAGTGGCAGAGTTCAGTTCGCTCTCTCT 140
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTYGlu 50
|||||
141 GAACCTCTAGGCAAGTTCAGTTCACAGATCTTACTACAAATATAGAG 190
51 IleValLeuThrGlnProValThrAlaAspLeuIleValLysLysG1 67
|||||
191 ATGTCTCTGGCAGCCGCTAACAGCTGACCTTATGTGTGAAGAAAAACA 240
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHis 84
|||||
241 AGTCCATTTTTCATATCATCTGATCTGTCGACAAATGTAAGCCCAT 290
84 euAsnValSerGlyIleProCySserValLeuLeuAlaAspValGluAsp 100
|||||
291 TAAATGAGCGGAAATTCATGCAGTGTCTGCTGCGACAGCGTAAGAT 340
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
341 CTATTCACAGCAGATTTCCACAGACAGTCAAGCCCGACCTCCCGC 390
117 aSerTYTYTYGluGlnTYrHisSerLeuAsnGluIleTYrSerTYrIleG 134
|||||
391 ATGTCATATGAGACAGTATCCTCATAATGAAATCTATCTTGATGAG 440
134 lUpheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||
441 AATTATTAACAGAGCAGCATCTGATATGCTTACAAATAATCCACATGGA 490
151 SerSerPheGluLysTYrProLeuTYrValLeuLysValSerGlyLysG1 167
|||||
491 TCCTCATTTGAGAGATACCCACTATATGTTTAAAGGTTCTGGAAAAA 540
167 uGlnThrAlaLysAsnAlaIleThrPLeaSPCySgIleHisAlaArgG 184
|||||
541 ACAAAACAGCAAAAATGCCATATGATGATGATGATGCAATCCATGCCAGG 590
184 lUThrPLeSerProAlaPheCySLeuThrPheIleGlyHis..... 197
591 AATGATCTCTCCGCTTCTGCTGTGCTCATAGGCCATATTAATCA 640
197 .....
641 TTCATAGGATATATAGGCAATATACCAATCTCCGAGGCTGTGGAAT 690
197 .....
691 CTAATGATCCGGTGCTTAATGTGAGCGTTATGATCACTACATGAGAAA 740
198 ..AsnArgMetThrPArgLysAsnArgSerPheTYrAlaAsnHisCys 213
|||||
741 AGAATCGAATGTGAGAAAGAACCGTCTTCTATGCAACATCATATGTC 790
214 lIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTYrCySgLuG1 230
|||||
791 ATCGAAGACGACCTGATATGAACTTTCCTCCAAACACAGTGATGAGGA 840

```



```

666 AAGACCCGCTACCAATGCTGGAATACCTGATTCGACAGACACCCCA 715
219 snSerAspHeValSerLyHisTrpCysgluInglYAlaSerSerSer 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
716 ACAGAAATTTT...GATGCTGGGTGGTGCACAACTGGAGCCTTACAGAC 762
236 SerCysSerGluThrTrpCysGlyLeuTrpProGluSerGluProGluVa 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
763 CCGTGCATAGACTACTGTGATCTGTCACAGATCTGAAAAAGAGAC 812
252 llyAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAla 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
813 CAAGCCCTGCTGATTTTATACGCAACAACTCTCTCCATCAAGCAT 862
269 yllSerMetHisSerTrpSerGlnHisIleValPheProTrpSerTrp 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
863 ACCTGACGATCCACTCATACACAGATGATCTACCTTATCTCAT 912
286 ThrArgSerLySerLyAspHisGluGluLeuSerLeuValAlaSerG1 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
913 GATTACAACTCCCGAGAACAACTGAGTGAATACCTGGCTAAGGC 962
302 uAlaValAlaIleAspLyThrSerLyAsnThrArgTrpThrHisG 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
963 TGGCGTAAAGAACTT...GCTACACTGTATGGCACAAGTACATACG 1009
319 lYHisGlySerGluThrLeuTrpLeuAlaProGlyGlyGlyAspAspTrp 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1010 GCCCAGAGGATACAAACATATCTGCTGCTGGGGGCTGTGATGACTGG 1059
336 lIeTrpAspLeuGlyIleLySTrpSerPheThr 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1060 GCTTATGACCAAGAAATATATCTCTTCCAC 1092

seq_name: /SIDSI/gcgdata/geneseq/emb1/NA196.DAT:AA12494
seq_documentation_block:
ID AA12494 standard; DNA; 1263 BP.
XX
AC AA12494;
XX
DT 12-FEB-1997 (first entry)
XX
DE Human pancreatic carboxypeptidase B encoding sequence.
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPPT;
KW mustard-ribonucleotide; antibody directed enzyme produg therapy;
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; F(ab')2; Fc; leader; human carboxypeptidase B; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1248
    /tag=a
    /product= preproenzyme
FT sig_peptide 1..39
    /tag=b
    /note="partial enzyme secretion leader sequence
    (pre-sequence); residues -108 to -96"
FT sig_peptide 40..324
    /tag=c
    /note="pro-sequence; residues -95 to -1"
FT mat_peptide 385..1245
    /tag=d
    /note="mature enzyme; encodes residues +1 to +307"
XX
XX WO9620011-A1.
XX
XX 04-JUL-1996.
XX
XX 21-DEC-1995. 95WO-GB02991.

```

```

XX
PR 16-AUG-1995; 95GB-0016810.
PR 23-DEC-1994; 94GB-0026192.
XX
XX (ZENEC ) ZENEC A LTD.
XX
PI Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
DR WPI: 1996-321650/32.
XX
PT P-PSDB; AAW06172.
XX
PT Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
XX anti-neoplastic prodrug
XX
PS Reference Example 14; Page 131; 182pp; English.
XX
XX A two-component system for anti-tumour therapy comprises a targeting
XX moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX prodrug. The system is based on antibody directed enzyme produg therapy
XX (ADPPT) using non-naturally occurring mutant forms of host enzymes,
XX pref. human pancreatic ribonuclease (HP-RNase), (see AA12478-83).
XX Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX be used. The present sequence encodes the native prepro-HCPB (see
XX AAW06172) contained in vector pIC11698.
XX
SQ Sequence 1263 BP; 343 A; 306 C; 302 G; 312 T; 0 other;

alignment_scores:
    Quality: 639.00 Length: 379
    Ratio: 2.526 Gaps: 7
Percent Similarity: 66.755 Percent Identity: 35.620

alignment_block:
US-09-980-881-2 x AA12494 ..

Align seg 1/1 to: AA12494 from: 1 to: 1263

6 LeuAlaValIleValProIleValLeuPheCysGluGlnHisValPheAl 22
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 CTCTTGCTTCGTGACTGCTGCGCCCTGGCATCTGCTCATGATGGTGGCA 53
22 a...PheGlnSerGlyGlnValIleValAlaLeuProArgTrpSerArg 38
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
54 GCACCTTGAAGCGAGAGAGTGTCCGTTAACCTGACATGAAATAC 103
38 lValGlnValIleGlnAsnLeuThrTrpTrpTrpGluIleValLeuTrp 54
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 ACATTAAACATAATCCGCGAGTTGGCCACGACGACCAATGTGACTCTGG 153
55 GlnProValThrAlaAspLeuIleValLysLysLysGlnValHisPheP 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 AAGCCAGATCTGTCCACCAACAAATCAACCTCAGTACAGTCACTCCG 203
71 eValAsnAlaSerAspValAlaAspValIleValHisLeuAsnValSerG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 TGTTAAGCAGAGATCTGTCACTGTGAGAAATGTTTAAAGCAGATG 253
88 lYlleProCysSerValLeuLeuAlaAspValGluAspLeuIleGln 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 AACATACATACAGTACTGATTAAGCAACCTGAGAAATGTGTGAGGCT 303
105 GlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTrpTrpG1 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 CAGTTGATAGCCGGGT.....CGTCAACAGCAGACAGATTATGA 344
121 uGlnTrpHisSerLeuAsnGluIleTrpSerTrpIleGluPheIleTrp 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 GAAGTACACAAGTGGGAAACGATTAAGAGCTTGGACTACACAACTGCCA 394

```

```

138 LuArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
139 CTEGAAATCCGACCCCTCATCTCTGAGCTTATCGGAACCAATTTGAG 444
145 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLys 171
146 GAGCGCCCTATTTCCTCTGTAAGTT...GGCAAGCTGGACAAATTA 491
171 GAGAAIleTrrPleAspCysGlyIleHisAlaArgGluThrPheSer 188
172 GCGCGCATTTTCATGAGCTGCTTCCATGACAGAGATGATTTCTC 541
188 roIAPheCysLeuThrPheLe... 195
542 CTGCAATTCCTGCAAGTGGTTGTAGAGAGCTGTTCTACTATGACGT 591
195 ..... 195
592 GAGATCCAGTGCAGAGCTTCTGCACAACTTAGACTTTATGCTCTGCC 641
196 .....GlyHis.....AsnArgMet 201
642 TGTGCTCAATATTGATGCTATCATCTACACCTGGACCAAGCCGATTTT 691
201 rPArGLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 217
692 GCGAAGAAAGATCGCTCCACCCATCTGATCTAGCTCATGGCCACAGAC 741
218 LeuAsnSerAsnPheValSerLysHisTrrPcysGluGluLysAlaSer 234
742 CCCAACAGAAATTTT...GATGCTGGTGGTGTGAATTTGAGACCTCTCG 788
234 rSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluPro 251
789 AAACCCCTGATGAAACCTTACTGCTGACCTGCCGACAGACTCGAAAGC 838
251 LuValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLys 267
839 AGACCAAGGCCCTGCTGATTTATCCGCAACAACTCTCTTCCATCAAG 888
268 AlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 284
889 GCATATCTGCATATCCATCCGCTACTCCCAAAATGATGATACCTTACT 938
284 rTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 301
939 ATATGCTTTCAAACCTCGGTGAGAACATCTGAGTTGAATGCCCTGCTA 988
301 exGluAlaValAlaArgAlaIleAspLysThrSerLysAsnThrArgTyr 317
989 AAGCTACTGTGAAGAACTT...GCCTCACTGCACGCGCAACAGTACACA 1035
318 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAsp 334
1036 TATGGCCCGGAGCTCAACAAATCTATCTGCTGGGGGCTGTGACGA 1085
334 pTrrPleTyrAspLeuGlyIleLysTyrSerPheThr 346
1086 CTGGGCTTATGACCAAGCATGATATCTTCAACC 1122
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:AAV41795
seq_documentation_block:
ID AAV41795 standard; DNA; 1263 BP.
XX
AC AAV41795:
XX
DT 20-NOV-1998 (first entry)
XX
DE Human pancreatic carboxypeptidase nucleotide sequence.
XX
KW ser: human; pancreatic carboxypeptidase B; insulin; protein sequencing;
prodng therapy.

```

```

XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 1..1248
FT /tag="a
FT /product="Pancreatic carboxypeptidase B"
FT /note="No start codon given"
XX
PN W09835988-A1.
XX
PD 20-AUG-1998.
XX
XX 10-FEB-1998; 98WO-GB00415.
XX
XX 29-OCT-1997; 97GB-0022727.
XX 14-FEB-1997; 97GB-0003104.
XX 18-OCT-1997; 97GB-0022003.
XX
XX (ZENE ) ZENEPA (ZENE ) ZENECA LTD.
XX
XX Edge MD:
XX
XX WPT: 1998-467168/40.
XX P-PSDB: AAM74476.
XX
XX New modified pro-domain of carboxy-peptidase B - enhances expression
XX of co-expressed proteins for production of recombinant
XX carboxy-peptidase or its fusions with antibodies, used, e.g. in
XX enzyme prodng therapy
XX
XX Example 1; Page 52-53; 83pp; English.
XX
XX The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
XX modified pro-domain of CPB on a separate gene to enhance recombinant
XX expression. This process can be used to produce recombinant CPB in
XX eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
XX antibody production and protein sequencing, while its fusions with
XX antibodies are useful in antibody-directed enzyme prodng therapy. The
XX CC modified pro-domain provide increased yields of recombinant CPB, possibly
XX CC by protecting the C-terminus against enzymatic degradation or by
XX CC increasing intracellular trafficking.
XX
XX Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;
XX
XX
XX alignment_scores:
XX Quality: 639.00 Length: 379
XX Ratio: 2.526 Gaps: 7
XX Percent Similarity: 66.755 Percent Identity: 35.620
XX
XX alignment_block:
XX US-09-980-881-2 x AAV41795 ..
XX
XX Align seg 1/1 to: AAV41795 from: 1 to: 1263
XX
XX 6 LuValAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAl 22
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 4 CTCTGGTCTGTGCTGACCTGGCCCTGGGACATCGTCATCATGAGTGGTGA 53
XX
XX 22 a...PheGlnSerGlyGlnValLeuAlaValLeuProArgThrSerArg 38
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 54 GCACCTTTGAAGCGAGAGAGTGTCTCGTTAACCTGACATGAGTGAATAATC 103
XX
XX 38 InValGlnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuThr 54
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 104 ACATTAACTAATCCGCGAGTTGGCCAGCAGACCCCAAGATTGACTTCTGG 153
XX
XX 55 GlnProValThrAlaAspLeuIleValLysLysGlnValHisPhePhe 71
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 154 AAGCCAGATTCTGTCAACAAATCAAACTCAGACTAGAGTTGACTTCTCG 203
XX
XX 71 eValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerG 88

```

```

204 TGTAAAGCAGAAATCTCTACTGTGGAGAAATGTTCTAAAGCAGAAATG 253
88 LylleProcySerValLeuLeuAlaAspValGluAspLeuIleGln 104
254 AACTACAAATCAAGTAGTCTAGTAAAGCAACCTGAGAAATGGGTGAGGCT 303
105 GluIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrGln 121
304 CAGTTGATAGCCGGT.....CTGCAACAGCAGCACTGTTATGA 344
121 uGIntYrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleArg 138
345 GAAGTACAAACAGTGGAGAAACGATAGAGCTGGATCAACACAGTCCCA 394
138 LuArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
395 CTGAGAAATCAGCCCTCATCTCTCGCAGTGTATACGAGACCACTTTGAG 444
155 LysTyrProLeuTyrValLeuLysValSerGlyLysGluIntHraIaly 171
445 GACGCGCCTATTACTCTGAAAGTT...GGCAACCTGGACAAATAA 491
171 sAnAlaIleTrpIleAspGlyIleHisAlaArgGluTrpIleSer 188
492 GCCTGCATTTTCATGAGCTGGTTCCATGACAGAGAGTGTCTC 541
188 roAlaPheCysLeuTrpPheIle..... 195
542 CTGATCTCTGCCAGTGTGTTGTAAGAGAGCTGTCTACTATGAGAGT 591
195 ..... 195
592 GAGATCCAGTAGCAGAGCTTTCGACAGTATAGACTTTATGTCCTGCC 641
196 ..... GlyHis ..... AsnArgMet 201
642 TGTGCTCAATATGATGCTACATCTACACCTGAGCAAGAGCCGATTTT 691
201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 217
692 GAGAAAGAGCTGCTCCACCATCTGATCTACTGATGGACAGAC 741
218 LeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSer 234
742 CCCAACAGAAATTTT...GATGCTGTGGTGTGAAATGAGCCCTCG 788
234 rSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProG 251
789 AAMCCCTGTGATGAACTTACTGTGAGCTGCCGACAGAGTCTGAAAGG 838
251 LuValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLys 267
839 AAMCAAGGCGCTGCGATTTTCATCCGACAAACATCTCTCCATCAAG 888
268 AlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 284
889 GCATATCTGCAATCCACTGCTACTCCCAATGATGATCTACCTTATCTC 938
284 rTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAla 301
939 ATATGCTTACAACTCGTGAGAACATGCTGAGTTAAATGCCCTGGCTA 988
301 erGluAlaValArgAlaIleAspLysThrSerLysAsnTrpArgTyrThr 317
989 AAGCTACTGGAAAGAACTT...GCCACACTGCAGGCGACCAAGTACACA 1035
318 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAs 334
1036 TATGGCCCGGAGCTACCAACATCTCTGCTGGGGGCTCTGACGA 1085
334 pTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 346

```

```

1086 CTGGGCTTATGACCAAGAAATCATCATATTCCTTCACC 1122
seq_name: /SID1/gcgdata/geneseq/geneseqn-emb1/NA196.DAT:AAAT42506
seq_documentation_block:
ID   AAT42506 standard; DNA; 1284 BP.
XX
AC   AAT42506;
XX
DT   17-FEB-1997 (first entry)
XX
DE   PROHCPB gene with PelB leader sequence.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodng therapy;
KW   anti-neoplastic; prodng; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
OS   Synthetic.
XX
FH   Key
FT   CDS
FT   sig_peptide
FT   sig_peptide
FT   sig_peptide
FT   mat_peptide
FT   mat_peptide
FT   mat_peptide
PN   WO9620011-A1.
PD   04-JUL-1996.
XX
PE   21-DEC-1995; 95MO-GB02991.
XX
PR   16-AUG-1995; 95GB-0016810.
PR   23-DEC-1994; 94GB-0026192.
XX
PA   (ZENEA ) ZENEA LTD.
PI   Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
PI   Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   P-PSDB; AAW06175.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti-neoplastic prodng
XX
PS   Reference Example 18; Page 140-142; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodng. The system is based on antibody directed enzyme prodng therapy
CC   (ADPPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence is a cloned pro-HCPB gene contained in
CC   plasmid pIC11738 and which can be expressed in E. coli.
XX
SQ   Sequence 1284 BP; 355 A; 315 C; 299 G; 315 T; 0 other;
alignment_scores:
Quality: 631.50      Length: 361
Ratio: 2.588         Gaps: 6

```

Percent Similarity: 67.590 Percent Identity: 35.734

alignment_block:

US-09-980-881-2 x AA142506 ..

Align seg 1/1 to: AA142506 from: 1 to: 1284

```

23 PheGlnSerGlnValLeuAlaLeuProArgThrSerArgLysIle 39
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 TTTGAGGCGAGGAGGTGTTCCGTGACGTTGAGATGAAATACACAT 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 LglnValLeuGlnAsnLeuThrThrThrGluIleValLeuThrLys 56
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 TAAACATATCCCGAGTTGGCCACGACGACCCAGATTGACTGTGGAGC 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 72
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 CAGATTCTGCACACAAATCAACCTCAGTACAGTTGACTTCGGTGT 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly 89
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 AAAGCAGAAATATCTGTCACTGTGGAGATGTTCTAAAGCAGATGAACT 284
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnIle 106
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 ACAATACAACTACTGATTAAGCAACCTGAGAAATGTGTGGAGGCTCAGT 334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 LeSerAsnAspThrValSerProArgAlaSerAlaSerLysGluGln 122
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
335 TTGATAGCCCGGTT.....CGTGCAACAGGACACACTTATGAGAAC 375
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 139
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
376 TACAAACAAGGGGAAAGATGAGGCTTGACATCAACATCCGCTCCTGA 425
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 156
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
426 GAATCCAGCCCTCATCTCTCCAGTGTATCGGACACCATTTGAGGAGC 475
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 yzProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLys 172
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 GGCCTATTTCCTCTCGAAGGTT...GGCAAGCTGGACAAATAAGCCT 522
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 AlaIleTrpLeuAspCysGlyIleHisAlaArgGluTrpLeuSerPro 189
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
523 GCCATTTCCTGAGCTGTGTTCATGCGAGAGTGTCTCTCTCTGCTGC 572
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 apPheCysLeuTrpPheIle..... 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
573 ATTCTGCCAGTGTGTGTAAGAGAGGCTGTCTACTATGACAGTGACAG 622
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 ..... 195
623 TCCAACTGACAGACTTTCGACAAAGTTAGACTTTATGTCCTGCTGTG 672
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 .....GlyHis.....AsnArgMetTrpArg 202
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
673 CTCATATATGATGGCTTACATCTACACCTGAGACCAAGCGGATTTGGAG 722
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
723 AAAGACTCCCTCCACCATCTAGCATCTAGCTGATGGACACAGACCCCA 772
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 snSerAsnPheValSerLysHisTrpCysGluGluValAspSerSer 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
773 ACAGAAATTTT...GATGCTGTGTGTGTGAATGTGAGACCTCTCGAAC 819
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 SerCysSerGluThrTyrCysGlyLeuTyrProGlnUserGluProGlu 252
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
820 CCCTGTGAGAACTTACTGTGTGACCTGCCGAGAGTGTGAAGAGAGAC 869
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAla 269
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

870 CAAGCCCGTGGCTGATTCATCCGCAACAAACCTCTTCATCAAGCAT 919
269 YrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 285
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
920 ATCTGACAAATCCTACTGCTACTCCCAATGATGATCACTTACTACTAT 969
286 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerG 302
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
970 GCTTACAAATCGGTGAGAAACATGCGAGTTGATGATCCCTGGCTAAC 1019
302 uAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHis 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1020 TACTGTAAAGAACTT...GCTTCACCTGCACGGCACCACCAATACATATG 1066
319 LysHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 335
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1067 GCCCGGAGCTTACAAACATCTATCTCTGCTGAGGCTCTGACGACTGG 1116
336 IleTyrAspLeuGlyIleLysTyrSerPheThr 346
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1117 GCTTATGACCAAGAAATCAGATATTCCTTCACC 1149
seq_name: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.AA17331
seq_documentation_block:
ID AA17331 standard; DNA; 2154 BP.
XX
AC AA17331;
XX
DT 02-JUN-1998 (first entry)
XX
DE PreproHCPB-linker-humanised 806.077 VH gene.
XX
KW Anti-CEA antibody; carcinoembryonic antigen; 806.077 Ab; cancer therapy;
  cancer diagnosis; complementarity determining region; ss.
XX
OS Synthetic.
XX
FH Key 1.2151 Location/Qualifiers
FT CDS /*tag= a
FT
XX
PN W09742329-A1.
XX
PD 13-NOV-1997.
XX
PF 29-APR-1997; 97WO-GB01165.
XX
PR 14-FEB-1997; 97GB-0003103.
XX
PR 04-MAY-1996; 96GB-0009405.
XX
PA (ZENEC) ZENECA LTD.
XX
PI Copley CG, Edge MD, Emery SC;
XX
DR WP1; 1997-558987/51.
XX
DR P-PSDB; AAM41414.
XX
PT Anti-carcinoembryonic antigen antibody 806.077 Ab - used for
  diagnosis and therapy of cancer.
XX
PS Example 75; Page 181-183; 208pp; English.
XX
CC This sequence encodes preproHCPB-linker-humanised 807.077 VH, which is
  an example of the antibody of the invention. The antibody is an anti-CEA
  (carcinoembryonic antigen) antibody (preferably 806.077 Ab). Host cells
  or transgenic organisms transformed with DNA encoding the antibody, are
  used to make the antibody or conjugate. The conjugate is used in a
  medicament suitable for intravenous administration. The conjugate can be
  used for cancer therapy, selectively killing tumour cells. The antibody
  can be used for in vivo or in vitro diagnosis of cancer.
XX
SO Sequence 2154 BP; 532 A; 601 C; 548 G; 473 T; 0 other;

```


XX Claim 1; Page 30; 36pp; English.
 CC The present sequence is a cDNA encoding novel human protein (NHP),
 CC carboxypeptidase-related protein. The carboxypeptidase-related
 CC proteins share structural similarity with animal proteases, and
 CC especially carboxypeptidase B or carboxypeptidase A. The
 CC carboxypeptidase-related protein cDNA is useful for the detection of
 CC mutant human carboxypeptidases, or inappropriately expressed human
 CC carboxypeptidases for the diagnosis of disease, and also as a
 CC therapeutic. It is also useful in drug screening, for generation of
 CC antibodies, as reagents in diagnostic assays, for the identification of
 CC other cellular gene products related to human carboxypeptidases, and as
 CC reagents in assays for screening for compounds that are useful for
 CC treating mental, biological or medical disorders and diseases. Nucleotide
 CC constructs encoding functional NHPs can also be used in gene therapy.
 XX

SQ Sequence 1311 BP; 391 A; 278 C; 284 G; 358 T; 0 other;

alignment_scores:
 Quality: 611.50 Length: 417
 Ratio: 2.427 Gaps: 12
 Percent Similarity: 60.432 Percent Identity: 33.813

alignment_block:

US-09-980-881-2 x AAD03837 ..

Align seg 1/1 to: AAD03837 from: 1 to: 1311

```

1 MetLysLeuGly.....SerLeuAlaValLeuValProIleValLeuPh 15
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
43 CTGCGCTTTTGGCGCTTTTGGAGATTCTGCACCG..... 81

15 eCySGluGlnHisValPhe.....AlaPheGlnSerGlyGlnValLeuA 30
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
82 .GGGCACAGCCACCTTATACAAACCGCTATGCTGGTGATAAAGTGTAA 130

30 LaIaLeuProArgThrSerArgGlnValGlnValLeuGlnAsnLeuThr 46
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
131 GATTATTTCCCAACAGAGAGAGACATATGCACGAAAGAAATATCC 180

47 ThrThrTyrGluIleValLeuThrGlnProValThrAlaAspLeuIleVa 63
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
181 TATCAACTTAAGGTGACCTGTGCGACGCCAGCATATCTCCATATGATC 230

63 Llys.....LysLysGlnValHisPhePheValAsnAlaSerAspValA 78
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
231 AGAGGAGACAGTACTGATGTCATATCCCCCAAAATGTTCCCGAGCC. 279

78 spAsnValLysAlaHisLeuAsnValSerGlyIleProCysSerValLeu 94
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
280 ....CTGTAGCCCTTCTTACAGAGAACCAACATCCAGTACAGAGTCCTC 324

95 LeuAlaAspValAlaLysAspLeuIleGlnGlnIleSerAsnAspThrVa 111
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
325 ATGAGAGATCTTCAGAAACACTGAGAGAGGAGAGAGCTTGACACCCA 374

111 LserProAlaGlnSerAlaSer.....TyrTyrGluGlnIleTyrHisSer 126
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
375 GAGAAACCGAGATCCCTCTGTGATATATATGAAATTATCACTCTCT 424

126 euAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAlp 142
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
425 TAGAAGAAATTCAAAATTGATGATCATCATGATTAATAACTCATCAGCC 474

143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTy 159
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
475 CTCATTCACATGTTCTCTATGTGAGAGATCATATGAGGAGAAATCTCTT 524

159 rValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpI 176
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
525 TATTTTAAAGCTG...GGCAGACGATCACGACTCAAAAGAGCTGTTTGA 571

```

```

176 leaSPcysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeu 192
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
572 TAGACTGTGGTATTCATGCAAGAAATGATGGTCTGCTGCTTTTGTACAG 621

193 TrpPheIle.....GlyHisAsnArgMetTrpArgLysAsnArg 195
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
622 TGGTTTAAAGAAAGCTCTTTCATCATATAGAGTACCCAGCATGAG 671

195 ..... 195

672 AAAAATTGAAATCATATATATTTATATCATAGCCGTGTTAAAGCTCG 721

196 .....GlyHisAsnArgMetTrpArgLysAsnArg 205
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
722 ATGGATACCATTTTACTGTGACCAATGATGATTTTGGAAAAACAGG 771

206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPh 222
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
772 TCAGAGAACTCAAGGTTCCGTCGCCGTGAGTGGATGCCAATGAAACTG 821

222 eValSerLysHisTrpCysGluGlnGluAlaSerSerSerSerCysSerG 239
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
822 GAAAGTGAAAG...TGGTGTATGAAAGAGGCTTCTATGACACCTTGATG 868

239 LuThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaVal 255
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
869 ACAATATCTGTGGCCCTTTCCAGAAATCTGAGCCGGAAGGAGGCTGTGA 918

256 AlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaIleTyrIleSer 272
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
919 GCTAACTTCCTTGAAACACAGAAAGACATTAAGGCTTATCTCTCTCT 968

272 ThrSerTyrSerGlnHisIleValPheProTyrSerTyrThrArgSerL 289
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
969 TCATGATATGCTCAGATGTTACTGTATCCCTATTTCTTCAAAATATGCA 1018

289 ysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 305
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1019 CAATTCCCAATTTTACATGATGTCGAAATCTGACAGCTTATMAAGCTGTG 1068

306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySe 322
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1069 GCACCTT...CAGTCAGTATACGGGGTACGATACAGATATGAGCCAGCC 1115

322 rGluThrLeuTyrLeuAlaProGlyGlyLysAspAspTrpIleTyrAspL 339
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1116 CACAACGTTGTATGTAGAGCTGTGAGCTCAGTCATGATGGGCTTACAAA 1165

339 euGlyIleLysTyrSerPheThr.Ser..... 347
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1166 ATGGAATACCTTATATGATTTGCTTTCGAACCTAGTCAGACACTGGATATTT 1215

348 .....AsnProProValGluLysLeu 354
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1216 GGATTTTACTCCAGAGATGCTCATCAAAACCCACTGTACGAAACATA 1264

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA1996.DAT.AAT35760
seq_documentation_block:
ID AAT35760 standard; cDNA; 927 BP.
XX
XX AAT35760;
XX
XX 09-NOV-1996 (first entry)
XX
XX Rat mature carboxypeptidase B cDNA.
XX
XX Carboxypeptidase-B; pro-enzyme; protease; insulin; ss.
XX
XX Rattus sp.
XX
XX FH Key Location/Qualifiers

```


This Page Blank (uspto)

OM of: US-09-980-881-2 to: Issued_Patents_NA:* out_format : pfs
Date: Sep 18, 2002 6:55 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+_p2n.model -DEV=x1h
-O=/cgn2_1/USPTO_spool/US09980881/rnmat_16092002_140129_10653/epc_query.fasta_1.1126
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=rm1 -CAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPEXT=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -TGAPOP=10.000 -TGAPEXT=0.500
-DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ps
-NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881.ecgn1.1.62 -NCP=6 -ICPUC=3 -LONGLOC
-DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPMX -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-2
Query length: 360
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 90.520000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation	..	5e-214	1272	1
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-869-057-1		1832.00	3976.37	8.5e-214	1272	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-07-649-591B-2		1832.00	3972.25	1.4e-213	1749	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-277-540-2		1832.00	3972.25	1.4e-213	1749	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-430-787A-2		1832.00	3972.25	1.4e-213	1749	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-696-139-1		640.50	1363.65	2.2e-68	1215	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-56		639.00	1361.85	3.6e-68	1263	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-71		631.50	1345.19	3.1e-67	1284	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-171-945-124		623.00	1319.87	7.9e-66	2154	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-782-760-5		604.00	1289.14	4.1e-64	927	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:PCr-US96-00995-5		604.00	1289.14	4.1e-64	927	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-696-139-3		576.00	1227.86	1.1e-60	921	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-67		561.00	1193.93	8.2e-59	999	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-54		561.00	1193.25	8.2e-59	1053	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-74		554.00	1177.84	6.4e-58	1059	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-860-882A-77		553.00	1175.64	8.5e-58	1059	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-171-945-112		543.00	1146.37	3.6e-56	1870	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-640-906-3		513.00	1085.82	8.6e-53	1251	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-395-936-3		513.00	1085.82	8.6e-53	1251	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-640-906-1		487.50	1029.87	1.1e-49	1257	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-395-936-1		487.50	1029.87	1.1e-49	1257	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-331-709-2		213.00	438.04	1.0e-16	591	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-998-416-125		168.50	342.29	2.2e-11	515	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-484-105-15		118.50	212.25	0.0004	2504	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-484-106-15		118.50	212.25	0.0004	2504	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-782-760-2		100.50	200.92	0.0017	285	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:PCr-US96-00995-2		100.50	200.92	0.0017	285	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-706-216-1		95.00	159.68	0.3317	2719	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-609-049A-29		95.00	151.08	0.9993	5285	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-170-996-29		95.00	143.88	2.51	4680	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-254-358-1		91.00	143.88	2.51	4680	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-475-391-1		91.00	143.88	2.51	4680	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-709-609-1		91.00	143.88	2.51	4680	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:PCr-US95-07178-1		91.00	143.88	2.51	4680	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-073-297-1		90.50	159.04	0.3600	1333	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-178-252-26		88.00	148.54	1.38	1965	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-689-974-2		87.50	150.30	1.10	1576	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-058-376-2		87.50	150.30	1.10	1576	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-07-915-203-1		86.50	136.73	6.30	3797	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-272-887-1		86.50	136.73	6.30	3797	1			
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-789-449-1		86.50	136.73	6.30	3797	1			

/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-609-049A-27 + 86.00 128.03 19.20 6831
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-09-170-996-27 + 86.00 128.03 19.20 6831
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-07-807-043B-9 + 85.00 132.27 11.15 4157
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-299-849B-9 + 85.00 132.27 11.15 4157
/cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-142-368A-9 - 85.00 132.27 11.15 4157

seq_name: /cgn2_6/pdata1/1/ina/5A.COMB.seq:US-08-869-057-1

seq_documentation_block:

Sequence 1, Application US/08869057
Patent No. 5985562
GENERAL INFORMATION:
APPLICANT: Morser, Michael J
TITLE OF INVENTION: Method of Detecting Thrombotic Disease
TITLE OF INVENTION: Risk
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Berlex Biosciences Legal Department
STREET: 15049 San Pablo Avenue
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94804-0099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,057
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Washlien, Wendy L.
REGISTRATION NUMBER: 36,301
REFERENCE/DOCKET NUMBER: 51509AUSM1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-262-5411
TELEFAX: 510-262-7095
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1272 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PUBLICATION INFORMATION:
AUTHORS: Eaton, Dan L.
AUTHORS: Malloy, Beth E.
AUTHORS: Tsai, Siao P.
AUTHORS: Hensel, William
AUTHORS: Drayna, Dennis
TITLE: Isolation, Molecular Cloning, and Partial
Characterization of a No. 5985562el Carboxypeptidase B
JOURNAL: J. Biol. Chem.
VOLUME: 266
ISSUE: 32
PAGES: 21833-21838
DATE: NO. 5985562 15-1991
US-08-869-057-1

alignment_scores:

Quality:	1832.00	Length:	415
Ratio: <td>5.103<td>Gaps:<td>3</td></td></td>	5.103 <td>Gaps:<td>3</td></td>	Gaps: <td>3</td>	3
Percent Similarity: <td>86.506<td>Percent Identity:<td>85.783</td></td></td>	86.506 <td>Percent Identity:<td>85.783</td></td>	Percent Identity: <td>85.783</td>	85.783

alignment_block:

US-09-980-881-2 x US-08-869-057-1 ..
Align seg 1/1 to: US-08-869-057-1 from: 1 to: 1272

```

1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
1 ATGAGAGCTTTCAGCCCTTCAGCTCTTACCAATGTTCTCTCTCTGCA 50
17 uG1H1stValPheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProA 34
51 GCAGCATCTCTCCGCTTTCAGAGGCGCAAGTCTTACCTGCTCTCTCTA 100
34 TgthSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyrGlu 50
101 GAACCTTAGGCAAGTTCAGATCTTACAGATCTTACCAACATATGAG 150
51 IleValLeuThrProValThrAlaAspLeuIleValLysLysG1 67
151 ATTCTTCTGCGCAGCCGCTAACAGCTCCTTATTGGAAGAAAAACA 200
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaH1st 84
201 AGTCATTTTCTTGTAAATGCATCTGATGTCAGCAATGTGAAGCCCAT 250
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
251 TAAATGTGAGCGGAAATTCATGAGTGTCTGCTGGCAGAGTGGAAGAT 300
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
301 CTATTTCACAGCAGATTTCACAGCAGACAGCAGCCCGAGCCCTCCG 350
117 aSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleG 134
351 ATGTAATGAGAGAGATATCATGAGTGTCTGCTGGCAGAGTGGAAGAT 400
134 LurheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
401 AATTAAATAGAGAGCATCTGATATGCTTACAAAAATCCACATGGA 450
151 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysG1 167
451 TCCTCATTTGAGAGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAGA 500
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
501 ACAAAAGCCCAAAAATGCCATATGATGATGCTGGAATCCATGCGCAG 550
184 LurTrpIleSerProAlaPheCysLeuThrPheIleGlyHis..... 197
551 AATGATCTCTCTGCTTCTGCTGTGTTGATAGGCCATATATACCTCAA 600
197 ..... 197
601 TTCTATGGGATATAGGCAATATACCAATCTTCCTGAGCGTTGTGATTT 650
197 ..... 197
651 CTATGTTATGCCGCTGTTAATGTGAGCGTTATGACTACTCATGAGAAA 700
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyrTrpAlaAsnHisCys 213
701 ACAAATCGAATGTGAGAAAGAACCGTCTTCTATGCGAACAATCATCTTC 750
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluG1 230
751 ATCGGAACAGACCTGAATGAGAACTTGTCTCCAAACAATGTTGTGAGGA 800
230 uGlyLysSerSerSerSerSerSerGluThrTyrCysGlyLeuTyrProG 247
801 AGGTGCATCTCAATCTCTCATGCTCGAAACCTACGTGACTTTATCTCG 850
247 LuSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
851 AGTCAGAACCAAGAACTGAAGCAGTGGCTAGTTTCTTGAGAAAGAAATATC 900

```

```

264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
901 AACCAAGATTAAAGCATATACATACAGATGATTCATACCTCCACGATATAGT 950
280 LpheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuS 297
951 GTTCCATATATCTATACAGCAAGTAAAGCAAGAAAGACATGAGGAATGCT 1000
297 eLrLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsn 313
1001 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATGAGAAACTAGTAAAAAT 1050
314 ThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrIleuAlaProG1 330
1051 ACCAGGTATACATATGGCCATGGCTCGAAGAACTTATACCTAGCTCTGG 1100
330 YGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer...Phe... 345
1101 AGGTGGGAGCAGATGATGATGATGATTTGGCATCAATATTCGTTACAA 1150
346 ..... 1150
1151 TTGAACCTTCAGATACGGGCAATACGATTTCTGCTGCCGAGCGTTAC 1200
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
1201 ATCAAAACCACCTGTAGAGAAAGCTTTGCCGCTGCTCTTAAAA 1243
seq_name: /cgn2.6/ptodata/1/ina/5A.COMB.seq:US-07-649-591B-2
seq_documentation_block:
; Sequence 2, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patln (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435
; FILING DATE: 19910201
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178

```

```

; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 41 to 106
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-07-649-591B-2

```

```

alignment_scores:
  Quality: 1832.00      Length: 415
  Ratio: 5.103          Gaps: 3
  Percent Similarity: 86.506  Percent Identity: 85.783

```

alignment_block:
US-09-980-881-2 x US-07-649-591B-2 ..

Align seg 1/1 to: US-07-649-591B-2 from: 1 to: 1749

```

1 MetIysLeuCySserLeuAlaValIleuValProIleValIleuPheCyseGI 17
  |||||
41 ATGAGAGCTTTCAGCCCTTCAGAGTCCCTGACCCCATGTTCTCTCTCTGAGA 90
17 uGIHISValAlPheAlaPheGInSerGlyGInValIleuAlaIleuProA 34
  |||||
91 GCACAGATCTCTCGCGCTTCAGAGTCCCATGTTCTACCTCTCTCTCTCTA 140
34 rGTHSerArgGInValGInValIleuGInAsnIleuThrThrThrTyrGlu 50
  |||||
141 GAACCTCTAGGCAAGTTCAGATCTTACAGATCTTACACAAATATGAG 190
51 ILeValLeuTrpGInProValThrAlaAspLeuIleValLysLysLysGI 67
  |||||
191 ATTCTCTCTGCGACGCCGTAAACAGCTGACCTTATTGTGAAGAAAAACA 240
67 nValHISpHePheValAsnAlaSerAspValAspAsnValLysAlaHISL 84
  |||||
241 AGTCATATTTTGTAAATGATCTGATGTCGACATGTGAAGCCCAT 290
84 euAsnValSerGlyIleProCySserValIleuLeuAlaAspValGluAsp 100
  |||||
291 TAAATGTGAGGGAATTCATGATGTCGTCTGCGACGAGCTGGAAT 340
101 LeuIleGInGInIleSerAsnAspThrValSerProArgAlaSerAl 117
  |||||
341 CTATTTCACAGCAGATTTCCACAGACAGACAGCCCGGAGCCTCCGC 390
117 aSerTyrTrpGluGInTyrHISerLeuAsnGluIleTyrSerTrpIleG 134
  |||||
391 ATCGTACTATGACAGTATCTCACTCAATAAGAAATCTATTCTTGAGAG 440
134 IupheIleThrGluArgHISProAspMetLeuThrLysIleHISIleGly 150
  |||||
441 AATTATTAACGAGAGCATCTGATATGCTTACAAAAATCCACATTGGA 490
151 SerSerPheGluLysTyrProLeuTyrValIleuLysValSerGlyLysGI 167
  |||||
491 TCCCATTTGAGAGTACCACCTCATGTATTAAAGTTTCTGGAAGAAGA 540
167 uGIHISValAlPheAlaPheGInSerGlyGInValIleuAlaIleuProA 184
  |||||
541 ACAAAACGACCAAAATGCCATATGATGACTGTGGAATTCATGCCAGAG 590
184 IuTrpIleSerProAlaPheCySserLeuTrpPheIleGlyHIS..... 197
  |||||
591 AATGATCTCTCTGCTTCTGCTGTGTTGTTCAATGAGCCATATATACCTGA 640

```

```

197 ..... 197
641 TTCTATGGGAATATAGGGCAATATACCAATCTCTGAGGCTTGTGAAATT 690
197 ..... 197
691 CTATGTTATGCCGGTGTATATGTGACGGTTATGACTACTCATATGAAAA 740
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHISCyS 213
  |||||
741 AGAATCGAATGTGAGAGAAAGAACCTTCTTCTATGACCAACATCATATGC 790
214 ILeGlyThrAspLeuAsnSerAsnPheValSerLysHISTrpCySGLuGI 230
  |||||
791 ATCGGACAGACCTGAAATAGCAACTTTCCTTCCAAACACTGGTGTGAGGA 840
230 uGIValaSerSerSerSerCySserGluThrTyrCySGLuLeuTyrProG 247
  |||||
841 AGGTGCAATCCAGTTCCTCATGCTCGGAACCTTACTGTGACTTATTCCTG 890
247 IuSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
  |||||
891 AGTCAGAACCAAGAGTGAAGGCGAGTGGCTAGTTCTTGAGAGAAATATTC 940
264 AsnGInIleLysAlaTyrIleSerMetHISerTyrSerGInHISIleVa 280
  |||||
941 AACCAATTTAAAGCATATACATGATGATCATCTCCACACATATATAGT 990
280 IPhProTyrSerTyrThrArgSerLysSerLysAspHISGluGluAs 297
  |||||
991 GTTTCATATTCCTTATACAGAAATGAAAGCAAAACCACTGAGGAATGT 1040
297 erLeuValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsn 313
  |||||
1041 CTTAGTATGCCAGTGAACAGTTCGTGATGAGAAAAATCTGTAATAAT 1090
314 ThrArgTyrThrHISGlyHISGlySerGluThrLeuTyrLeuAlaProGI 330
  |||||
1091 ACCAGGTATACATGATGCCATGCTCAGAAACCTTATACCTGATCTCCG 1140
330 yGIgLyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe... 345
  |||||
1141 AGGTGGGAGCATGTGATCTATGATTTGGGCATCAAAATTTGTTTACAA 1190
346 .....Th 346
1191 TTGAACCTCGAGATACGGGCAATACGATCTTGTCTGCGGAGCGTTAC 1240
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
  |||||
1241 ATCAAAACCCACCTGTAGAGAAAGCTTTTGCCCTGCTCTTAAAAA 1283

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-277-540-2

seq_documentation_block:

```

; Sequence 2, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Dreyer, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; City: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-277-540-2

alignment_scores:
Quality: 1832.00      length: 415
Ratio: 5.103          Gaps: 3
Percent Similarity: 86.506      Percent Identity: 85.783

alignment_block:
US-09-980-881-2 x US-08-277-540-2 ..
Align seg 1/1 to: US-08-277-540-2 from: 1 to: 1749

1 MetLysLeuGysSerLeuAlaValLeuValProLLeuValLeuPheCysG1 17
|||||
41 ATGAAGCTTTGCAAGCTTGGCAAGTCTTGACCCATGCTTCTCTTCGTGA 90
|||||
17 uGlnHsValPheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuPro 34
|||||
91 GCAGCATGCTTGGCTTCAAGTGTACAGAGTCCCAAGTTCTAGCTCTCTCA 140
|||||
34 rGHisSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyrGlu 50
|||||
141 GAACCTCTAGGCAAGTCAAGTCTACAGAACTTACTACAAACATATGAG 190
|||||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
|||||
191 ATTGTCCTCTGGCAGCCGGTAACAGCTGACCTTATTTGTGAAGAAAAACA 240
|||||
67 nValHsPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
|||||
241 AGTCCATTTTGTGTAATGCATCTGATGTCGACATGTGMAAGCCCATTT 290
|||||
84 euAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
|||||
291 TAAATGTGAGCGAATTCATGACAGTCTCTGCGACAGCGTGAAGAT 340
|||||
101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
341 CTTATTCAACACACAGATTTCACACGACACAGTCCAGCCCGGCGCTCCGC 390
|||||
117 aSerTyrTyrGlnGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleG 134
|||||
391 ATCGTACTATGAACAGTATCACTCACTAAATGAAATCTTTTGTGATNG 440
|||||
134 IupHeIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150

```

```

|||||
441 AATTATTAAGTGAAGGCAATCCTGATATGCTTACAAAAATCCATTGCA 490
|||||
151 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysG1 167
|||||
491 TCCCTATTGTGAAGTACCCACTGATGTTTAAAGGTTCTGTGAAAAAGA 540
|||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
|||||
541 ACAAAACAGCCAAAAATGCCATATGATGATGAGTGGAAATCCATGCCAG 590
|||||
184 IupTrpIleSerProAlaPheCysLeuTrpPheIleGlnHis..... 197
|||||
591 AATGATCTCTCCGCTTTCGCTTGCTTGCTTCAAGGCCATATATACTCAA 640
|||||
197 ..... 197
641 TTCTATGGATTAATAGGCAATATACCAATCTCCTGAGCCTTGATTT 690
|||||
197 ..... 197
691 CTATGTTATCCGGTGTATATGACGCTTATGACTACTATGGAATA 740
|||||
198 ..AsnArgMetTyrArgLysAsnArgSerPheTyrAlaAsnAsnHisCys 213
|||||
741 AGAATCAATGTGGAGAAAGACCGTCTTCTATGCGAACAATCATTTGC 790
|||||
214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluG1 230
|||||
791 ATCGGAACAGACCTGAATAGAACTTGTCCAAACACAGTGTGACGA 840
|||||
230 uGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrProG 247
|||||
841 AGGTGATCCAGTTCCTCATGCTCGGAACCTTGTGACTTATTCCTG 890
|||||
247 IuSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
|||||
891 AGTCAACACGAAAGTGAAGGCGAGTGTAGTTCTTGAGAAATAATATAC 940
|||||
264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
|||||
941 AACCAATTAAGCAATACATACAGATGCATTTCACTATCCACCAATATG 990
|||||
280 IuPheProTyrSerTyrThrArgSerLysSerLysAspHisGluLys 297
|||||
991 GTTTCATATTCCTATACAGAAATGAAGCAAAAGCATGAGAACTGT 1040
|||||
297 erLeuValAlaSerGluAlaValArgAlaIleAspLysHisSerLysAsn 313
|||||
1041 CTTTACTACCAAGTGAAGCAAGTTCGTGCTATGAGAAAACTATGTAATAAT 1090
|||||
314 ThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProG1 330
|||||
1091 ACCAGTATACACATGGCCATGGCTCAGAAACCTTATACCTTACCTCTGG 1140
|||||
330 yGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer..Phe... 345
|||||
1141 AGGTGGGAGCATGTGATCTATGATTTGGCATCAAAATATTCGTTACAA 1190
|||||
346 .....Th 346
1191 TTGACATTCGAGATACGGGACATACGGAATTCCTGCTCCGAGCGTTAC 1240
|||||
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
|||||
1241 ATCAAAACCACCTGTAGAGAAAGCTTTTCCGCTGTCTATAAA 1283
|||||
seq_name: /cgn2_6/ptodata1/lna/5a_comb.seq:US-08-430-787A-2
seq_documentation_block:
; Sequence 2, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:

```

```

APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-8881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-430-787A-2

alignment_scores:
Quality: 1832.00      Length: 415
Ratio: 5.103          Gaps: 3
Percent Similarity: 86.506      Percent Identity: 85.783

alignment_block:
US-09-980-881-2 x US-08-430-787A-2 ..

Align seg 1/1 to: US-08-430-787A-2 from: 1 to: 1749

1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCySgl 17
|||||
41 ATGAAGCTTTCACAGCTTCAGAGCTTCACCAATGCTCTCTCTGGA 90
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProA 34
|||||
91 GCACAGATGCTTCGCGTTTCAGAGTGGCCAGATCTAGCTCTCTCTA 140
|||||
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyGlu 50
|||||
141 GAACCTCTAGGCAAGTCAAGTCTACAGAACTTACTACAAACATATGAG 190
|||||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGlu 67
|||||
191 ATTGTTCTGTGCGACCGGTAAACAGCTGACCTTATTGTGGAAGAAAAACA 240

```

```

67 nValHisPhePheValAsnAlaSerAspValAlaAspAsnValLysAlaHisL 84
|||||
241 AGTCATTTTTCGTAATGCAATCTGATGTCAGACAATGTGAAGCCCAT 290
|||||
84 euAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
|||||
291 TAAATGTAGCGGAATTCATGCACTGCTGCTGCGACAGCCGTGAAGAT 340
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
341 CTTATTCAACAGCAGATTTCCACAGACACAGTACACCCCGAGCCTCGC 390
|||||
117 aSerTyTrpGluGlnTyHisSerLeuAsnGluIleTySerTrpIleG 134
|||||
391 ATCGTACTATGACAGTATCACTCACTAAATGAATCTATCTTGATAG 440
|||||
134 lPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||
441 AATTTATACGTAGAGCATCTCGATATGCTTACAAAATCCACATTCGA 490
|||||
151 SerSerPheGluLysTyProLeuTyValLeuLysValSerGlyLysGlu 167
|||||
491 TCCTATTTGAGAACTACCCACTATGTTTAAAGTTTCGGAAGAA 540
|||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
|||||
541 ACAACACGCAAAATGCCATATGATGACTGTGAATCCATGCCAGAG 590
|||||
184 lUTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
|||||
591 AATGATCTCTGCTGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 640
|||||
197 ..... 197
641 TTCATGGGATTAATGGGCATATACCAATCCCGAGGCTTGATGATT 690
|||||
197 ..... 197
691 CTATGTATGCCGCGTGTAAATGTGAGCGGTATGACTACTCATGAGAAA 740
|||||
198 ..AsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnHisLys 213
|||||
741 AGAATCGAATGTGAGAAAGAACCGTCTTCTTATGCAACACATCATGTC 790
|||||
214 lIleGlyThrAspLeuAsnSerAsnPheValSerLysHisIleTrpCysGlu 230
|||||
791 ATCGGACAGACCTGATAGAGAACTTCTTCCAAACACTGCTGAGAGA 840
|||||
230 uGlyAlaSerSerSerCysSerGluThrTyTrpCysGlyLeuTyProG 247
|||||
841 AGGTGCATTCAGATTCCTCATGCTCGAAGAACCTAGTGAATTTATCCTG 890
|||||
247 lUserGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
|||||
891 AGTCAGAACCAAGAGTGAAGCGAGTGTGCTAGTTCTTGAAGAAATATC 940
|||||
264 AsnGlnIleLysAlaTyrlIleSerMetHisSerTySerGlnHisIleVa 280
|||||
941 AACCGATTTAAAGCATATCATCATGATCATCATCTCCACGACATATAGT 990
|||||
280 lPheProTySerTyThrArgSerLysSerLysAspHisGluLeuS 297
|||||
991 GTTTCATATTCCTATACAGCAAGTAAAGCAAAACCATGAGAACTGT 1040
|||||
297 erLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsn 313
|||||
1041 CTCTAGTACCAAGTGAAGCAGTTCGTGCTATTTGAAGAAATCTAGTAAAT 1090
|||||
314 ThrArgTyThrHisGlyHisGlySerGluThrLeuTyTrpLeuAlaProG 330
|||||
1091 ACCAGGTATACAGATGCGCATGCGCTCAGAAACCTATATCTAGCTCCGCG 1140
|||||
330 yGlyGlyAspAspTrpIleTyAspLeuGlyLysLysTySer.Phe... 345

```



```

251 luvalysalavalaserpheluarlgarganlleasnglllelys 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
839 AGACCAAGCCGCGGATTCATCGCAACAACCTCTTCATCAAG 888
268 AlATrIleSerMetHISerTYrSerGlnHisIleValPheProtyrse 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
889 GCATATGTGCATCATCCCTGCTACTCCCAATGATGATCTTACCTTACTC 938
284 rTyrrHrArSerLysSerLysAspHisGluGluLeuSerLeuValAlas 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
939 ATATGCTTACAACTCGTGAGAACATGCTGACTGAATGCCCTGCTGCTA 988
301 erGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrrHr 317
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
989 AACCTACTGGAAGAACTT...GCCCTACTGCACGCGCAACCACTACACA 1035
318 HIsGlyHIsGlySerGluThrLeuTyrrLeuAlaProGlyGlyAspAs 334
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 TATGGCCCGGAGCTACACATCTATCCCTGCTGCGGCTCTGACGA 1085
334 pTrpIleTyrrAspLeuGlyIleLysTyrrSerPheThr 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 CTGGCGCTTATGACCAAGCAATCATCTCTCTCACC 1122

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-860-882A-71

seq_documentation_block:

; Sequence 71, Application US/08860882A

; Patent No. 5985281

; GENERAL INFORMATION:

; APPLICANT: TAYLORSON, CHRISTOPHER JOHN

; APPLICANT: EGGELTE, HENDRIKUS JOHANNES

; APPLICANT: TARRAGONA-FIOL, ANTONIO

; APPLICANT: RABIN, BRIAN ROBERT

; APPLICANT: BOYLE, FRANCIS THOMAS

; APPLICANT: HENNAM, JOHN FREDERICK

; APPLICANT: BLAKELY, DAVID CHARLES

; APPLICANT: MARSHAM, PETER ROBERT

; APPLICANT: HEATON, DAVID WILLIAM

; APPLICANT: DAVIES, DAVID HOW

; TITLE OF INVENTION: CHEMICAL COMPOUNDS

; NUMBER OF SEQUENCES: 77

; CORRESPONDENCE ADDRESS:

; ADDRESS: PILLSBURY, MADISON & SUTRO

; STREET: 1100 NEW YORK AVENUE, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/860, 882A

; FILING DATE: JUNE 23, 1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: DONALD J. BIRD

; REGISTRATION NUMBER: 25,323

; REFERENCE/DOCKET NUMBER: 9901/238653

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3027

; TELEFAX: (202) 822-0944

; TELE: 6174627 CUSH

; INFORMATION FOR SEQ ID NO: 71:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1284 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-860-882A-71

alignment_scores:

Quality:	631.50	Length:	361
Ratio:	2.588	Gaps:	6
Percent Similarity:	67.590	Percent Identity:	35.734

alignment_block:

US-09-980-881-2 x US-08-860-882A-71 ..

Align seg 1/1 to: US-08-860-882A-71 from: 1 to: 1284

```

23 PheGlnSerGlyGlnValLeuAlaIleuProArgThrSerArgGlnVa 39
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 TTTCAGAGCGAGAAAGTTCGCTTACGTTAAGCTTGAAGATGAATCAT 134
39 lGlnValLeuGlnAsnLeuThrThrTyrrGluIleValLeuTropIn 56
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 TACATATATCCGCGAGTTGGCCACGACGACCAGATGACTTCTGGAAC 184
56 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 72
||:|||||:|||||:|||||:|||||:|||||:|||||:
185 CAGATTCTGTACACAAATCAAACTCAGATGACTTCTGCTTCCGCTT 234
73 AsnAlaSerAspValAlaAspAsnValLysAlaHisLeuAsnValSerGly 89
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 AAAGCAAGAGATACTCTCACTGTGAGAAATGTTTAAAGCAGAAATGAA 284
89 eProCySerValLeuLeuAlaAspValGluAspLeuIleGlnGlnIn 106
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 ACAATACAGAGTACTGTATTAAGCAACTGAGAAATGTGTGAGAGCTCAG 334
106 leSerAsnAspThrValSerProArgAlaSerTyrrTyrrGluGln 122
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
335 TTGATAGCCGGTT.....CGTCCACAGACAGACACTTATGAGAAC 375
123 TyrrHisSerLeuAsnGluIleTyrrSerTrpIleGluPheIleThrGluAr 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
376 TACACAAAGTGGGAAACGATGAGAGCTTGGACATCAAAAGTCCCACTGA 425
139 gHisProAspMetLeuThrLysLysHisIleGlySerPheGluLysTr 156
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
426 GAATCCAGCCCTCATCTCTGCAAGTGTATCGGACCACTTTGAGAGGAC 475
156 yrProLeuTyrrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 172
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 GCGCTATTTCACCTCGAAGGTT...GCCAAAGCTGGACAAATTAAGCCT 522
173 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
523 GCCATTTCATGAGCTGTGTTCATGCCAGAGAGTGGATTTCTCTCCGCG 572
189 aPheCysLeuTrpPheIle..... 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
573 ATTCTGCCAGTGTGTGTGAAGAGAGGCTGTCTACTATGACGCTGAGA 622
195 ..... 195
623 TCCAGTGCACAGACTTCTGCACAGTGTAGACTTTATGCTGCTGCTG 672
196 .....GlyHis.....AsnArgMetTrpAr 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
673 CTCATATGATGAGCTTACATCTACACCTGACCAAGACGAGTGGAG 722
202 glyAsnArgSerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeuA 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
723 AAAGACTCCGCTCACCATATCTAGATCTAGCTATGGACACAGACCCCA 772
219 snSerAsnPheValSerLysHisTrpCysGluGluValAlaSerSerSer 235
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
773 ACAGAAATTTT...GATCGTGTGGTGTGAATGAGACCTCTCGAAGC 819
236 SerCysSerGluThrTyrrCysGlyLeuTyrrProGlnSerGluProGluVa 252

```

```
|||||
820 CCTGTGATGAACCTTACTGTGACCTGCCGACAGTCTGAAAGGAGAC 869
252 LlysAlaValAlaSerPheLeuArgAsnIleasnGlnIleLysAla 269
      |||||
870 CAAGCCCTGGCTGATTCATCCGCAACAACCTCTCCATCAAGCAT 919
269 yTlleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 285
      |||||
920 ATCTGAACATCTCACTGCTACTCCAAATGATGATCACTACCTTACTCATAT 969
286 ThrArgSerLysSerLysAspHisGlnGluLeuSerLeuValAlaSerGln 302
      |||||
970 GCTTACAAACCTGGTGAGAACATCTGAGTTGATGACCTGGCTGAAGC 1019
302 uAlaValAlaGlnIleAspLysThrSerLysAsnThrArgTyrThrHisG 319
      |||||
1020 TACTGTGAANAACCTT...GCCTCACCTGCAGGACCAAGATACACATATG 1066
319 LysHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTTP 335
      |||||
1067 GCCCGGAGAGCTCAACAATCATCTATCTGCTGCTGGGCGCTGACAGACTGG 1116
336 IleTyrAspLeuGlyIleLysTyrSerPheThr 346
      |||||
1117 GCTTATGACCAAGAAATCATATCTCTTGCACC 1149
seq_name: /cgn2_6/prodata/1/lna/5B_COMB.seq:us-09-171-945-124

seq_documentation_block:
; Sequence 124, Application US/09171945
; Patent No. 6277399
; GENERAL INFORMATION:
; APPLICANT: Emery, Stephen
; APPLICANT: Copley, Clive Graham
; APPLICANT: Edge, Michael Derek
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said
; FILE REFERENCE: Antibody, and their Therapeutic Use in an Adept System
; CURRENT FILING DATE: 1998-10-29
; PRIOR FILING DATE: 1997-04-29
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9703103.3
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 124
; LENGTH: 2154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-124

alignment_scores:
      Quality: 623.00      Length: 382
      Ratio: 2.462      Gaps: 7
      Percent Similarity: 66.230      Percent Identity: 34.555

alignment_block:
US-09-980-881-2 x US-09-171-945-124  ..
Align seg 1/1  to: US-09-171-945-124  from: 1  to: 2154

      3 LeucySerLeuAlaValLeuValProIleValLeuPheCysGlnGlnI 19
      ::::|  |||||  ::::|  ::::|
      1 ATGTTGGACATCTTGTTGTTGTTGTTGACCTGGCCTGGCATCTGCATCA 50
      19 sValPheAla...PheGlnSerGlyGlnValLeuAlaIleuProArgT 35
```

```
51 TGTGCTGAGACACTTTTGAAGCGAGACAGGTGTCCGTGTTAACTGTAAG 100
35 hrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThyTyrGluIle 51
      ::::|  |||||  ::::|  |||||
101 ATGAATATCACATTTATACATATCCCGAGTGTGGCCAGACACACCATGATT 150
52 ValLeuTyrGlnProValThrAlaAspLeuIleValLysLysGlnVal 68
      |||||
151 GACTCTCGAAGCCAGATTCTGTACACAAATCAAACTCAACAGTACAGT 200
68 LHisPhePheValAsnAlaSerPValAspAsnValLysAlaHisLeu 85
      |||||
201 TGACTCTCGGTGTTAAAGCAGAAATATCTGATCTGAGATGATGTTCTTA 250
85 snValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
      |||||
251 AGCAGATGAATGATACAAATACAAAGTACTGATTAAGCAACCTGCAAAATGTG 300
102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAla 118
      |||||
301 GTGAGGCTCAGATTGATAGCGGGT.....CGTGCAACAGACA 341
118 rTyTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPileGluP 135
      |||||
342 CAGTTATGAGAAATGACAAAGTGGAACAGATAGAGCTTGAGCTGACTCAAC 391
135 heIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySer 151
      |||||
392 AAGTGGCCACATGAGAAATCCAGCCCTCATCTCCGAGTGTATTCGGAACC 441
152 SerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluG 168
      |||||
442 ACATTTGAGGAGCGGCTATTACCTCCGGAAGTT...GGCAAAAGCTGG 488
168 nThrValAsnAlaIleTyrPileAspCysGlyIleHisAlaArgGluT 185
      |||||
489 ACAAAATAGACCTGCCATTTTCATGACTGTGGTTTCATGCCAGAGAGT 538
185 rPileSerProAlaPheCysLeuTyrPheIle..... 195
      |||||
539 GGATTTCTCTGCAATTCCTGCCAGTGTGTTGTAAGAGAGCTGTCTGACC 588
195 ..... 195
589 TATGAGCTGAGATCCAAAGTGACAGAGCTTCTGCAGAAATTAGACTTTTA 638
196 .....GlyHis.....A 198
639 TGTCTGCTGCTGTGCTCAATATTGAGGCTACATCTACACTGGACCAAGA 688
198 snArgMetTyrArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 214
      |||||
689 GCCGATTTTGGAGAAAGACTGCTCCACCCATACGAGATCTAGCTCATTT 738
215 GlyThrAspLeuAsnSerAsnPheValSerLysHisTyrCysGlnGluG 231
      |||||
739 GGCACAGAACCCCAAGAAATTTT...GATGCTGTTGGTGTGAATATGG 785
231 yAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGluS 248
      |||||
786 AGCCCTCTGAAACCCCTGTGATGAACCTTACTGTGACCTGCCGACAGGT 835
248 erGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
      |||||
836 CTGAAGAAGAGACCAAGGCCCTGGCTGATTTTCATCCGCAACAACCTCTCT 885
265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 281
      |||||
886 TCCATCAAGGATATCTGATCAATCCATCTGATCCCAAAATGATGATCTA 935
281 eProTyrSerTyrThrArgSerLysSerLysAspHisGlnGluLeuSerL 298
      |||||
```

```

936 CCCTTACTCATATGCTTACAACATCGGTGAGACACATGCTGAGTTGAATG 985
298 euValAlaSerGluAlaValAlaArgAlaIleAspLysThrSerLysAsnThr 314
986 CCGGTGCTAAGACTAGTGTGAAGAACTT...GCCTCACGACGACGGCACC 1032
315 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGly1 331
1033 AAGTACACAAATATGCCCCGGAGGACTACAAACAATCTATCTCTGCTGGGAC 1082
331 yGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 346
1083 TTCTAAGAGACTGGGCTTATGACCAAGCAATCAGATATTCCTTCACC 1128

seq_name: /cgn2.6/pldata/1/lna/5B.COMB.seq:US-08-782-760-5

seq_documentation_block:
; Sequence 5, Application US/08782760
; Patent No. 5948668
; GENERAL INFORMATION:
; APPLICANT: Hartman, Jacob
; APPLICANT: Fulga, Nelta
; APPLICANT: Mendelovitch, Simona
; APPLICANT: Gorecki, Marian
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,760
; FILING DATE: 13-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/778,233
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..927
; US-08-782-760-5

alignment_scores:
Quality: 604.00 Length: 269
Ratio: 3.374 Gaps: 5
Percent Similarity: 66.543 Percent Identity: 44.610

```

```

alignment_block:
US-09-980-881-2 x US-08-782-760-5
Align seg 1/1 to: US-08-782-760-5 from: 1 to: 927

115 AlAserAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 131
1 GCAAGTTCACCAAGTTGCGACCTGATATATCCAGACTTGTCTCACAAGCG 50
131 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 148
51 GTGGATTCAACAAAGTTGCGACCTGATATATCCAGACTTGTCTCACAAGCG 100
148 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
101 TCATTGGAACACACATTGGAAGACGTACATGTATGCTTCCCAAGATT... 147
165 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 181
148 GGTAAACACTAGACCGAATAGCCTGCCATCTGCATGATGTTGTTCCA 197
181 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle..... 195
198 TGCACAGAGTGGATTCTCTGCAATTCGTGACGTGTTGTGAGAGAG 247
195 ..... 195
248 CTGTCCGTACCTATATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297
196 ..... GlyHis..... 197
298 CTGGATTTCTATGTTCTGCTGTGTCACAACTGATGATGCTATGCTACAC 347
198 .....AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnA 211
348 CTGAGACTAAGACAGACATGTGGAGAAAAACCCCTTACTATAGCTGGGA 397
211 snHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
398 GTTCCTGCTGGGTGTAGACCCACACAGAAATTT...AATGTGCTGG 444
228 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 244
445 TGTGAAGTGGAGCTTCTGGAATCCCTGCTCTGGAACCTTACTGTGGACC 494
244 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgA 261
495 AGCCCAAGAGTCTGAAAAAGACAAAGCCCTGGCAGATTTCATCCGCA 544
261 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 277
545 ACAACTCTCCACACATCAAGGCTTACTGACATCCATCACATCAATCAGAG 594
278 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisI 294
595 ATGATGCTCTTACCTTACTCTATGACTACAAACTGCTGGAACACTATGA 644
294 uGluLeuSerLeuValAlaSerGluAlaValAlaArgAlaIleAspLys 311
645 GGAATTGTAATGCCCTGCTGAAGGTGGCGCAAAAGAGCTT...GCCAC 691
311 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
692 TGCATGGCACCAAGTACATATAGCCCCAGAGCTACAAACATCTATCT 741
328 AlaProGlyLysGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSe 344
742 GCTGCTGGGGGATGTGACGACTGTCTTATGATCAAGGAATCAATATATTC 791
344 rPheThr 346
792 CTTTACC 798

```


ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,139
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,258
FILING DATE: 16-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-8681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..921
US-08-696-139-3

alignment_scores:
Quality: 576.00 Length: 264
Ratio: 3.310 Gaps: 5
Percent Similarity: 65.909 Percent Identity: 43.939

alignment_block:
US-09-980-881-2 x US-08-696-139-3 ..

Align seg 1/1 to: US-08-696-139-3 from: 1 to: 921

```
120 TyTGLuGlnIuThHISerLeuAnGluIleTySerTrpIleGluPheI 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 TATGAGAAAGTACAAACAACTGGAAAGATGAGAGCTTGACTAAGCAAGT 65
136 eThrGluArgHisProAspMetLeuThrIleHisIleGlySerSerp 153
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
66 CACCAGTGAATAATCCAGACCTCATCTCGCACAGCCATCGGAATCAT 115
153 heGluIuSTyPrProLeuTyValLeuIuysValSerGlyIuSgluGlnThr 169
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 TTTTACGAAACAATATATACCTCTCAAGGTT...GGCAAACTGAGCA 162
170 AAlaIysAsnAlaIleTrpIleAspCysGlyIleHisAlaIArgGluTrpI 186
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 AATAAGCGCTGCCATTTTCATGAGCTGTGTTCCATGCCAGAGAAATGAT 212
186 eSerProAlaPheCysLeuTrpPheIle..... 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 TTCCCATGCAATTTTCCAGTGGTTGTGAGAGAGAGCTGTTTCACCTATG 262
195 ..... 195
263 GATATAGAGTCAATGACAGAAATTCCTCAACAAGCTAGACTTTTATGTC 312
196 .....GlyHis.....AsnAr 199
313 TTGCCTGTGCTCAATATGATGCTTACATCTACACTGAGCAAGAACCG 362
199 gHeCTrPArgIysAsnArGSerPheTherIAlaAsnAsnHisCysIleGlyT 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
363 AATGTGAGAAAGACCCCTCTACCAATGTCTGAACTACCTGCATTTGCCA 412
216 hrAspIeAsnSerAsnPheValSerIySHISrPcysGluGluIAla 232
|||||:|||||:|||||:|||||:|||||:|||||:|||||
413 CAGACCCCAACAGAAATTTT...GATGCTGGGTGTGTGACAACTGGAGCC 459
233 SerSerSerSerCysSerGluTrpTrpCysGlyLeuTyProGluSerG 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||
460 TCTACAGACCCCTCGATGAGACTTACTGTGATCTGCTGAGAGTGTGA 509
249 uProGluValIysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnI 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||
510 AAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACACCTCTCCCA 559
266 IeIysAlaTyIleSerMetHisSerTySerGlnHisIleValPhePro 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||
560 TCAAGCATACCTGACATCATCCTACACAGATGATACCTTACCT 609
283 TySerTyThrArgSerIysSerIysAspHisGluIuLeuSerIeUva 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||
610 TATTCCTATGATTACAAACTCCCGAGACAAATGCTGATGAATAACT 659
299 IAlaSerGluAlaValArgAlaIleAspIySHISrPcysGluIArgT 316
|||||:|||||:|||||:|||||:|||||:|||||:|||||
660 GGCTAAGGCTCGCTGAAGAAGACTT...GCTACACTGTATGACCAAGT 706
316 yThrHisGlyHisGlySerGluThrLeuTyIleAlaProGlyGlyGly 332
|||||:|||||:|||||:|||||:|||||:|||||:|||||
707 ACAATACGGCCCGAGAGCTACCAACAATCTATCTGCTGCTGGGGCTCT 756
333 AspAspTrpIleTyAspIleGlyIleTySerPheThr 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||
757 GATGACTGGGCTTATGACCAAGAAATCAATATCTTCCAC 798
```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-860-882A-67

seq_documentation_block:
Sequence 67, Application US/08860882A
Patent No. 5985281
GENERAL INFORMATION:
APPLICANT: TAYLORSON, CHRISTOPHER JOHN
APPLICANT: EGGELTE, HENDRIKUS JOHANNES
APPLICANT: TARRAGONA-PIOL, ANTONIO
APPLICANT: RABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNEM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3027
 TELEFAX: (202) 822-0944
 TELEX: 6174627 CUSH
 INFORMATION FOR SEQ ID NO: 67:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 999 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-860-882A-67

alignment_scores:

Quality: 561.00 Length: 269
 Ratio: 3.117 Gaps: 5
 Percent Similarity: 66.914 Percent Identity: 40.892

alignment_block:

US-09-980-881-2 x US-08-860-882A-67 ..

Align seg 1/1 to: US-08-860-882A-67 from: 1 to: 999

```

115 AlaserAlaserTYRGIUGINTYRHISerLeuansgluiletyrse 131
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCACTGCTACTCTTACGAGAACTACACACAGTGGAAACGATAGAGCC 116
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 rTPIleGlupheIleThgluRhIsProaspMetLeuThLysIleH 148
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TTGACTCAACAAGTCGCCACTGAGATTCACGCCCTCATCTCCGACGTG 166
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 IsIleGlySerSerPhegluLulysTYRProleuTYRValIleLysValSer 164
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTTTCGGAACACATTGAGGAGCGCGCTATTACCTCTGAAAGTT... 213
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 GlyIlegluInThraIalysAsnAlaIleTPRileaspCysglYleH 181
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GGCAAAAGCTGCACAAATAAGCCTGCCATTTCATGACGCTGTGTTCCA 263
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 salargluIleTPRileSerProalaphecYsLeuTPRphle..... 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TGCCAGAGAGTGGATTCTCTGCATTTCTCCAGTGGTTGTAGAGAGG 313
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 ..... 195
314 CTGTGTGACTATGAGAGTGAAGTCAAGTGAAGCTTCTCGACAG 363
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 ..... 197
364 TTAGACTTTATGTCCCTGCTGCTCAATATTGATGGCTACATCTACAC 413
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 ..... AsnArgMetIleTPRlyAsnArgSerPheTYRAlaAsn 211
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 CTGACCAAGAGCCGATTGGAGAAAGACTCGCTCCACCATCTGAT 463
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 snIleCysIleGIYThraPleuansSerAsnPhelaSerIlyshIsTPR 227
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 CTAGCTGATTTGACAGACCCCAACAGAAATTT...GATGCTGGTTGG 510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 CyaglIuGlulYalaserSerSerCysSerCysgluThryrCysglYle 244
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511 TGTGAATTTGAGCTCTCGAACCCTGTGATGAATCACTACTGTGACC 560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 uTYRProgluSerGluprogluValIyAlaValAlaserPheleuArg 261
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 TGCCGCAAGTCTGAAGAGACCAAGCCCTGGCTGATTTATCCGCA 610
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 rGAsnIleAsnGlnIleYalATyrIleSerMetHisSerTYRserGln 277
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611 ACAAACTCTTCCATCAGGCAATCTGACAAATCCACTGACTCTCCAA 660
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
278 HisIleValPheProTYRserTYRThraSerIySerIyAspHISgl 294
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 ATGATGATCTACCTTACTCATATGCTTACAAACTGGTGAAGCAATGC 710
  
```

```

294 uGluIleSerIleuValAlaserGlulAlaValArgAlaIleaspIlyshrs 311
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
711 TGAGTTGATGCTGCTGCTAAAGCTACTGTGAAGAACTT...GCCATC 757
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 eRlyAsnThraTYRThryshIsGlyshIsSerGlulThleuTYRleu 327
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
758 TGCAGGCAACCACTACATATGCCCCGGAGCTACACACATCTATCT 807
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 AlaprogIyglYAspAspTPRileTYRAspleuGlYleIlyTYRse 344
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
808 GCTGCTGGGGCTCTGAGCACTGGCGCTTATGACCAAGCAATCATATTC 857
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
344 rPheThr 346
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
858 CTCACCC 864
  
```

seq_name: /cgn2_6/prodata/1/lna/5b_COMB.seq:US-08-860-882A-64

seq_documentation_block:

; Sequence 64, Application US/08860882A
 ; Patent No. 5985281

GENERAL INFORMATION:

APPLICANT: TAYLORSON, CHRISTOPHER JOHN
 APPLICANT: EGSEITE, HENDRIKUS JOHANNES
 APPLICANT: TARRAGONA-FIOL, ANTONIO
 APPLICANT: RABIN, BRIAN ROBERT
 APPLICANT: BOYLE, FRANCIS THOMAS
 APPLICANT: HENNAM, JOHN FREDERICK
 APPLICANT: BLAKELY, DAVID CHARLES
 APPLICANT: MARSHAM, PETER ROBERT
 APPLICANT: HEATON, DAVID WILLIAM
 APPLICANT: DAVIES, DAVID HOW
 TITLE OF INVENTION: CHEMICAL COMPOUNDS
 NUMBER OF SEQUENCES: 77
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PILSBURY, MADISON & SUTRO
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk
 COMPUTER: IBM compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/860,882A
 FILING DATE: JUNE 23, 1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: DONALD J. BIRD
 REGISTRATION NUMBER: 25,323

REFERENCE/DOCKET NUMBER: 9901/238653

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3027
 TELEFAX: (202) 822-0944

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1053 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-860-882A-64

alignment_scores:

Quality: 561.00 Length: 269
 Ratio: 3.117 Gaps: 5
 Percent Similarity: 66.914 Percent Identity: 40.892

alignment_block:

US-09-980-881-2 x US-08-860-882A-64 ..

Align seg 1/1 to: US-08-860-882A-64 from: 1 to: 1053

```

115 A1aSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSe 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCACACTGGTCACCTTTACGAGAGTACAAACAGTGGGAAACGATAGAGGC 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 rTrrilegUpheleThrGluArgHisProAspMetLeuThrLysIleH 148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TTGGACTCAACAAGTCGCCACTGAGAAATCCAGCCCTCATCTCCGCGAGT 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 Is1leGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTAATCGAACAACATTGTGAGGAGCGGCTATTACCTCCGTAAGGTT... 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 G1LyLysG1uGlnThrAlaLysAsnAlaIleTrrPleAspCysGlyIleH 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GCGAAGAGTCGACAAATAAGCCCTGCCATTTTCATGAGCTGCTTCCCA 263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 sAlaArgGluTrrPleSerProAlaPheCysLeuTrrPhele..... 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TGCACAGAGAGTGATTTCTCTGCATTTCTGCCAGTGTGTTGTAAGAGG 313
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 ..... 195
314 CTGTTCGTACATAAGAGCTGAGATCCAAAGTACAGAGCTTCGACAAAG 363
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 ..... 197
364 TTAGACTTTATGTCTCGCTGCTGCCATATTGATGAGCTACTACATCAGC 413
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 ..... 197
414 CTGACCAAGAGCCGATTTTGGAGAAAGACTCGCTCCACCACCTAGTGAT 463
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 sHnHsCysAlleGlyThrAlaLeuAsnSerAsnPheValSerLysHsIstr 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 CTAGCTGGCTTGGACAGACCCCAAGCAAGAAATTT...GATCGTGTGG 510
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 CysGluGluGlyAlaSerSerSerSerSerSerSerGluTrrTyrCysGly 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511 TGTGAATGTGAGAGCCCTCGAAACCCCTGTGATGAAGAACTTACTGTGAGC 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 uTyrProGluSerGluProGluValAlaValAlaSerPheLeuArg 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 TGCSCCAGAGTCTGAAGAGAGCAAGCCCTGCTGATTCATCCGCA 610
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611 ACNAACTCTTCATCAAGGCAATGTGCAATCCATCCATCCGACCCCA 660
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
278 HsIleValPheProTyrSerTyrTrrArgSerLysSerLysAspHisG1 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 ATGATGATCTACCTTACTCATATGCTTACAAACACGCGAGAACATATCC 710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
294 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysTrs 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
711 TGAGTTGAATGCCCTGCTAAAGCTACTGTGAAGAACTT...GCCTAC 757
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 eTlysAsnThrArgTrrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
758 TGCACGGGACCAAGTACATATGGCCCGGAGTACAAACATCATATCTCT 807
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 AlaProGluGlyLysAspArgTrrIleTyrAspLeuGlyIleLysTyrSe 344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
808 GCTGTGGGGGCTCTGACGAGCTGGGCTTATGACCAAGATCAGATATTC 857
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
344 rPheThr 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
858 CTTCAC 864
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-860-882A-74

seq_documentation_block:

Sequence 74, Application US/08860882A

Patent No. 5985281

GENERAL INFORMATION:

```

APPLICANT: TAYLORSON, CHRISTOPHER JOHN
APPLICANT: EGGELE, HENDRIKUS JOHANNES
APPLICANT: FARRAGONA-FIOL, ANTONIO
APPLICANT: RABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNAM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
APPLICANT: DAVIES, DAVID HOW

```

TITLE OF INVENTION: CHEMICAL COMPOUNDS

NUMBER OF SEQUENCES: 77

CORRESPONDENCE ADDRESS:

ADDRESSEE: PILLSBURY, MADISON & SUTRO

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/860, 882A

FILING DATE: JUNE 23, 1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: DONALD J. BIRD

REGISTRATION NUMBER: 25,323

REFERENCE/DOCKET NUMBER: 9901/238653

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3027

TELEFAX: (202) 822-0944

TELEX: 6174627 CUSH

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 1059 bases

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-860-882A-74

alignment_scores:

Quality:	554.00	Length:	269
Ratio:	3.095 <td>Gaps:</td> <td>5</td>	Gaps:	5
Percent Similarity:	66.543	Percent Identity:	40.520

alignment_block:

US-09-980-881-2 x US-08-860-882A-74 ..

Align seg 1/1 to: US-08-860-882A-74 from: 1 to: 1059

```

115 A1aSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSe 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCACACTGGTCACCTTTACGAGAGTACAAACAGTGGGAAACGATAGAGGC 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 rTrrilegUpheleThrGluArgHisProAspMetLeuThrLysIleH 148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TTGGACTCAACAAGTCGCCACTGAGAAATCCAGCCCTCATCTCCGCGAGT 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 Is1leGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTAATCGAACAACATTGTGAGGAGCGGCTATTACCTCCGTAAGGTT... 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```
414 CTGGACCAAGACCGATTTTGGAAAGACTCGCTCCACCCATACGAT 463
      ::::::::::::::::::::::::::::
211 snH1aCys11egLYthraSPleuAnSerAsnPheValSerLYsH1aSTrp 227
      ::::::::::::::::::::::::::::
464 CTAGCTGCATTGGCACAGACCCCAAGAAATTTT...CATGCTGTTGG 510
      ::::::::::::::::::::::::::::
228 CysglugluGlyAlaSerSerSerSerCysSergluThrTYrCysglYle 244
      ::::::::::::::::::::::::::::
511 TGTGAAATTTGAGACCTCTCGAAACCCCTGTGATGAAACTTACTGTGACC 560
      ::::::::::::::::::::::::::::
244 UTyrProgluSergluProgluValLYsAlaValAlaSerPheleuArga 261
      ::::::::::::::::::::::::::::
561 TGCCCGCAGAGTCTGAAAGAGAGACCAAGGCCCTGGCTGATTCATCCGCA 610
      ::::::::::::::::::::::::::::
261 rGAsn11leAnGln11leLYsAlaTYr11leSerMeth11leSerTYrSerGln 277
      ::::::::::::::::::::::::::::
611 ACAAACTCTCTTCATCAAGGCATATCTGACAATCCACTCTACTCCAA 660
      ::::::::::::::::::::::::::::
278 H1s1leValPheProLYrSerTYrThraSerLYsSerLYsAspH1sG1 294
      ::::::::::::::::::::::::::::
661 ATGATGATCTACCTTACTCATATGCTTACAAACTCGTGAGAACAAATGC 710
      ::::::::::::::::::::::::::::
294 uGluLeuSerLeuValAlaSerGluAlaValArgAla11leAspLYsThrs 311
      ::::::::::::::::::::::::::::
711 TGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAGAACTT...GCCCTCAC 757
      ::::::::::::::::::::::::::::
311 eLYsAsnThraArgTYrThraH1sG1YH1sG1YSerGluThrLeuTYrLeu 327
      ::::::::::::::::::::::::::::
758 TGCACGGCACAAGTACACATATGGCCCGGAGCTACAAACATCTATCCT 807
      ::::::::::::::::::::::::::::
328 AlaProglYglYAlaSPasPTrip1eTYraSPleuGly11leLYsTYrSe 344
      ::::::::::::::::::::::::::::
808 GCTGCTGGGGGCTCTCGGCACTGGGCTTATGACCAAGAAATCAGATATTC 857
      ::::::::::::::::::::::::::::
344 rPheThr 346
      ::::::::::
858 CTTCAAC 864
```

OM of: US-09-980-881-2 to: EST:* out_format : pfs

Date: Sep 18, 2002 5:41 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+gzn_model -DEV=xlh
-o=/gcn2.1/USPO_spool/US09980881/runat_16092002_140128_10623/app-query.fasta_1.1126
-DB=EST -QFMT=fastap -SUFEXT=st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -TGAPOP=10.000 -TGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODAL=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09980881 @CN1.1.3434
-MCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLIPX -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-980-881-2
Query length: 360
Database: EST:*
Database sequences: 13736207
Database length: 1841457050
Search time (sec): 3115.820000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb.htc:AK004045	1534.50	3066.48	1.5e-161	1446	AK004045 Mus musculus 18 days
gb.est1:AL536069	1324.50	2648.24	2.9e-138	889	AL536069 AL536069 lrt1_F1013.F1013
gb.est2:BG616629	1034.00	2063.27	1.1e-105	788	BG616629 602646064F1 NIH_MGC_76
gb.est2:BT144644	1015.50	2025.01	1.5e-103	853	BT144644 602909932F1 NCI_CGAP_1
gb.est2:BG614456	996.50	1968.94	1.7e-101	738	BG614456 602614618F1 NIH_MGC_76
gb.est2:BG656051	983.00	1960.95	5.6e-100	747	BG656051 602585652F1 NIH_MGC_76
gb.est2:BG618239	975.00	1945.15	4.2e-99	725	BG618239 602645967F1 NIH_MGC_76
gb.est2:BG656281	965.00	1923.10	7.2e-98	852	BG656281 60258573F1 NIH_MGC_76
gb.est2:BT332659	957.50	1905.85	6.5e-97	1022	BT332659 602984408F1 NCI_CGAP_1
gb.est1:AV646979	949.00	1892.51	3.6e-96	735	AV646979 AV646979 GIC Homo sapi
gb.est2:BG618813	927.50	1848.88	9.8e-94	750	BG618813 602646166F1 NIH_MGC_76
gb.est2:BG656100	902.00	1797.45	7.1e-91	747	BG656100 60258370F1 NIH_MGC_76
gb.est2:BG656240	898.00	1790.28	1.8e-90	691	BG656240 60258370F1 NIH_MGC_76
gb.est1:AI048167	884.50	1762.49	6.3e-89	724	AI048167 ud71812.Y1 Sugano
gb.est1:AV693037	868.00	1731.69	3.3e-87	583	AV693037 AV693037 GIC Homo sapi
gb.est1:AV658930	847.00	1709.12	5.9e-86	602	AV658930 AV658930 GIC Homo sapi
gb.est1:AV694058	848.00	1692.99	4.7e-85	601	AV694058 AV694058 GIC Homo sapi
gb.est2:BG619111	837.50	1667.23	1.3e-83	749	BG619111 602645566F1 NIH_MGC_76
gb.est1:AV651709	818.50	1629.87	1.5e-81	688	AV651709 AV651709 GIC Homo sapi
gb.est2:BG657264	815.50	1621.98	4.2e-81	806	BG657264 602585945F1 NIH_MGC_76
gb.est1:AT529872	806.50	1606.11	3.3e-80	661	AT529872 u183811.Y1 Sugano
gb.est1:AT182113	806.00	1606.43	3.1e-80	580	AT182113 ud73110.Y1 Sugano
gb.est1:AA6968307	806.00	1606.43	3.1e-80	589	AA6968307 uc71h07.Y1 Sugano
gb.est1:AM950708	803.50	1601.11	6.2e-80	603	AM950708 EST362770 MAGE resequ
gb.est1:AT536068	799.50	1593.59	1.6e-79	575	AT536068 AL536068 lrt1_F1013.F1013
gb.est2:NG8450	773.50	1541.90	1.2e-76	537	NG8450 za28c09.X1 Soares fetal
gb.est1:AT532676	767.00	1527.09	1.6e-75	558	AT532676 mj24c06.Y1 Soares
gb.est1:AT574117	766.00	1522.34	7.9e-76	511	AT574117 u194809.Y1 Sugano
gb.est1:AT528146	745.00	1450.21	4.9e-75	511	AT528146 u194804.Y1 Sugano
gb.est1:AV681868	742.00	1450.06	3.4e-73	462	AV681868 AV681868 GMB Homo sapi
gb.est2:BG216526	741.50	1472.10	9.5e-73	843	BG216526 RST36106 Atherys RAGE
gb.est2:WB8434	731.00	1448.03	5.8e-72	455	WB8434 zhs9f04.X1 Soares fetal
gb.est1:AV658178	728.50	1449.35	1.8e-71	623	AV658178 AV658178 GIC Homo sapi
gb.est1:AV658128	718.00	1448.31	2.4e-70	615	AV658128 AV658128 GIC Homo sapi
gb.est1:AT574064	707.00	1406.20	4.4e-69	610	AT574064 u167911.Y1 Sugano
gb.est1:BE235838	694.00	1381.28	1.1e-67	447	BE235838 1434408 MARC lptg Sus s
gb.est1:AT574057	688.00	1370.45	4.3e-67	584	AT574057 u167904.Y1 Sugano
gb.est1:AM413090	685.50	1360.84	1.5e-66	723	AM413090 uq52c03.X1 Sugano

gb.est1:AT525929 - 672.50 1333.84 4.8e-65 772 | AT525929 u183811.X1 Sugano m
gb.est2:BF384322 + 663.50 1316.94 4.2e-64 692 | BF384322 602046670F1 NCI_CGA
gb.htc:AK003061 + 661.00 1303.51 2.3e-63 1429 | AK003061 Mus musculus adult
gb.est1:AB050299 - 651.50 1296.02 6.1e-63 520 | AB050299 AB050299 lambda zap
gb.est1:AT55397 - 646.50 1281.74 3.8e-62 747 | AT55397 u194804.X1 Sugano m

seq_name: gb.htc:AK004045

seq_documentation_block:

LOCUS AK004045 1446 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:1110032P04:carboxypeptidase B2 (plasma),
full insert sequence.

ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
clone:1110032P04.
MUS musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

REFERENCE

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL

MEDLINE 99279253
PubMed 10349636

REFERENCE

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

MEDLINE 20499374
PubMed 11042159

REFERENCE

Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohtsuka, E., Matsuura, S.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multiplexillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

MEDLINE 20530913
PubMed 11076861

REFERENCE

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL

MEDLINE 5 (bases 1 to 1446)

REFERENCE

Aach, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arawaka, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirotsu, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numata, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Schmitt, L., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Sohma, H., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, T., Toyota, T., Yamamura, T., Yamana, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

TITLE

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

JOURNAL


```

1089 ACCAGTACACACAGCGAGTGGCTCAGAAAGTTATATCTAGCTCTGG 1138
330 YGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer.Phe... 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1139 AGTTCGTGACGATGATGATCTATGATTTGGCATCAAAATATTCGTTACAA 1188
346 .....Th 346
1189 TTGAGCTCCGAGATACAGGAGATACGAGATTCCTGCTGCCAGAGATAC 1238
346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1239 ATCAAAACCACTTGTGCGAAGAGCTTTGGCCGCATCTCTAATA 1281

seq_name: gb_est1:AL536069

seq_documentation_block:
LOCUS AL536069 889 bp mRNA linear EST 13-FEB-2001
DEFINITION AL536069 lrl_FL013_FBrl Homo sapiens CDNA clone CS0DF022YE21 5
prime, mRNA sequence.
ACCESSION AL536069
VERSION AL536069.1 GI:12799562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 889)
AUTHORS Li,W.B., Gruber,C., Jesssee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..889
location/Qualifiers
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="CS0DF022YE21"
  /dev_stage="Prl.FL013_FBrl"
  /dev_stage="pooled tissue from post conception fetuses (20
  week, 24 week and 26 week)"
  /lab_host="DH10B"
  /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
  cDNA was primed with a NotI-oligo(dt) primer. Five prime
  end enriched, double-stranded cDNA was digested with Not I
  and cloned into the Not I and Eco RV sites of the
  pCMVSPORT 6 vector. Library was constructed by Life
  Technologies. Contact : Feng Liang Life Technologies, a
  division of Invitrogen 9800 Medical Center Drive Rockville
  , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
  fliang@lifetech.com URL :
  http://fulllength.invitrogen.com"

BASE COUNT 255 a 186 c 192 g 235 t 1 others
ORIGIN

alignment_scores:
  Quality: 1324.50 Length: 294
  Ratio: 5.174 Gaps: 1
Percent Similarity: 87.075 Percent Identity: 86.395

alignment_block:
us-09-980-881-2 x AL536069 ..
Align seg 1/1 to: AL536069 from: 1 to: 889

5 SerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPhe 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3 AGCCTTGACGTCCTTGATCCCATTTGCTCTCTTGAGCAGCAGTCCTT 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
21 eAlaPheGlnSerGlyGlnValLeuAlaValProArgThrSerArg 38

```

```

53 CGCGTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCCATGAAACCTTAGGC 102
38 lInValGlnValLeuGlnAsnLeuThrThrThrTyrGlnIleValLeuTrp 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
103 AAGTTCAGGTTCTACAGAACTTCTACTACAACTATAGAGATTCTTCTCGG 152
55 GluProValThrAlaAspLeuIleValLysLysLysGlnValHisPhePh 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 CAGCCGGTACAGCTGACCTTATGTGAAAGAAAAAACAAGTCATTTT 202
71 eValAsnAlaSerAspValAlaAspAsnValLysAlaHisLeuAsnValSer 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 TGTAAATGATCTGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCG 252
88 lYleProCysSerValLeuLeuAlaAspValGluAspLeuIleGln 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
253 GAATTCATGACAGTGTCTTGGCGACGACGTGAAGATCTTATACACAG 302
105 GlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrG 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
303 CAGATTTCCAAACGACACAGTCAGCCCGAGCCTCCGATCGTACTATGA 352
121 uGlnTrpHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThr 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
353 ACAGTATCACTCACTAAATGAATCTATCTTGATAGATTTATATACGT 402
138 lUArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
403 AGAGCATCTCTATATGCTTACAAAATTCACATTGATCTCATTTTGA 452
155 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAla 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
453 AAGTACCCACCTCTATGTTTAAAGTTCTGTGAAAAGAACAGCAGCCAA 502
171 sAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSer 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
503 AAATGCCATATGATGATGACTGTGGAATCCATCCAGAAATGATCTCTC 552
188 rAlaPheCysLeuTrpPheIleGlyHis..... 197
553 CTGCTTTCCTGCTTGTGTTCAATAGGCCATATATCAATTCATGGGATA 602
197 ..... 197
603 ATAGGCAATATACCAATCTCCTGAGCGTTTGATTTCTATGATAGCC 652
198 .....AsnArgMet 201
653 GGTGTTATATGTGATGTTATGACTACTCATGAAAAAAGAAATCGAATGT 702
201 rArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 217
703 GGAGAAAGACCGTTCTTCTATGCGAACAAATCATCTGCTTGAACAGAC 752
218 LeuAsnSerAsnPheValSerLysHisTrpCysGluGlnValAlaSerSe 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
753 CTGAATPAGAACTTTGCTTCCAAACACTGTGTGAGGAAGTCATCCAG 802
234 rSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProG 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
803 TTCTCATGCTCTCGAAACCTACTGTGACTTATCTCTGATCAGAACAG 852
251 lValLysAlaValAlaSerPheLeuArg 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
853 AAGTAAGGCACTGGCTAGTTCTTGAGAGA 884

seq_name: gb_est2:BG618629

seq_documentation_block:
LOCUS BG618629 788 bp mRNA linear EST 18-APR-2001
DEFINITION 602646064F1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4767974 5',
mRNA sequence.

```

```

ACCESSION      BG618629
VERSION        BG618629.1
KEYWORDS       GI:13670000
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 788)
AUTHORS        NIH-MGC http://mgc.ncl.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: c9abs-remail.nih.gov
               Tissue Procurement: CLONTECH Laboratories, Inc.
               cDNA Library Preparation: CLONTECH Laboratories, Inc.
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLM1630 row: 0 column: 15
               High quality sequence stop: 760.
               Location/Qualifiers
                 1..788
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:4767974"
                   /clone_11b="NIH_MGC_76"
                   /lab_host="DH10B (T1 phage-resistant)"
                   /note="Organ: Liver; Vector: pMDR-LIB (Clontech); Site: 1;
                   sfil1 (ggccatcggcc); Site: 2; sfil1 (ggccatcggcc); 5' and
                   3' adaptors were used in cloning as follows: 5' adaptor
                   sequence: 5'-CACGGCATATAGGCC-3' and 3' adaptor sequence:
                   5'-ATTCTAGAGCGGCGGCGGCGCATATG-dT(30)BN-3' (where B = A,
                   C, or G and N = A, C, G, or T). Average insert size 1.85
                   kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
                   by PCR. This library was enriched for full-length clones
                   and was constructed by Clontech Laboratories (Palo Alto,
                   CA). Note: this is a NIH_MGC Library."

BASE COUNT      222 a 157 c 169 g 240 t
ORIGIN

alignment_scores:
  Quality: 1034.00      Length: 241
  Ratio: 5.094          Gaps: 1
  Percent Similarity: 84.232      Percent Identity: 83.402

alignment_block:
  US-09-980-881-2 x BG618629 ..

Align seg 1/1 to: BG618629 from: 1 to: 788

1 MettyleucysSerleuAlaValleuValProIleValleuPheCysG1 17
  |||||||
66 ATAAAGCTTGGACGCTTGCAGTCCTGTACCATGTTCTCTTCGTGA 115
  |||||||
17 uGlnHlsvalPheAlaPheGlnSerGlyGlnValleuAlaAlaLeuPro 34
  |||||||
116 GCAGCATGCTTGGCTTGCAGTGGCCAAAGTTTACGTGCTCTCTCA 165
  |||||||
34 rGthSerArgGlnValGlnValleuGlnAsnLeuThrThrTrpGlu 50
  |||||||
166 GAACGCTCTGGCAAGTTCTACAGATCTTACTACACATATGAG 215
  |||||||
51 lIleValleuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
  |||||||
216 ATTGTTCTCTGGAGCGGCTTACAGCTGACTATGTGAAGAAAAACA 265
  |||||||
67 nValHlsPhePheValAlaAlaSerArgValAspAsnValLysAlaHls 84
  |||||||
266 AGTCATTTTGTAAATGACATCTGATGTGCAACATGTGAAGCCCAAT 315

```

```

84 eUAsnValSerGlyIleProCysSerValleuLeuAlaAspValGluasp 100
  |||||||
316 TAAATGTACGCGGAATTCATGACAGTGTCTTCTGGCAGATGTGAAGAT 365
  |||||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProAlaAlaSerAl 117
  |||||||
366 CTATATCAACAGCAGATTTCCACAGACACAGTACGCCGCCAGCTCCGC 415
  |||||||
117 aSerTyTrpGlnGlnTrpHisSerLeuAsnGluIleTySerTrpIleG 134
  |||||||
416 ATCTGACTATGACAGATGATCTACCTAAATGAATCTATTTCTGGATAG 465
  |||||||
134 lUpheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
  |||||||
466 AATTATTAACAGACAGCATCTGATATGCTTACAAAAATCCACATTTGCA 515
  |||||||
151 SerSerPheGluLysTyTrpLeuTyValleuLysValSerGlyLysG1 167
  |||||||
516 TCCTCATTTGAGAGTACCACCTGATGTTTAAAGSTTCTGGAAAGA 565
  |||||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
  |||||||
566 ACAAGCAGCCAAATGCAATGATGATGATGACTGGAATCCATGCCAGAG 615
  |||||||
184 lUTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
  |||||||
616 AATGATCTCTCCGCTTCTGCTGTGTTCAAGGCCATTAACACTCA 665
  |||||||
197 ..... 197
666 TTCTATGAGGATATAGGCAATATTAACAATCTCTGAGGCTTGATG 715
  |||||||
197 ..... 197
716 TCATATGATCCAGGCTTAATGTTGAGGTTATGACTACTATGCA 765
  |||||||
198 .....AsnArgMetTrpArgLys 203
  |||||||
766 AAAGCAATCATTTGTGAGAAAG 788

seq_name: gb_est2:BI144644

seq_documentation_block:
  LOCUS      BI144644      853 bp      mRNA      linear      EST 05-JUL-2001
  DEFINITION 60290932P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5051358 5',
  mRNA sequence.
  ACCESSION  BI144644
  VERSION    BI144644.1 GI:14604645
  KEYWORDS   EST.
  SOURCE      house mouse.
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
  REFERENCE  1 (bases 1 to 853)
  TITLE      NIH-MGC http://mgc.ncl.nih.gov/.
  JOURNAL    National Institutes of Health, Mammalian Gene Collection (MGC)
  COMMENT    Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: c9abs-remail.nih.gov
               Tissue Procurement: Jeffrey E. Green, M.D.
               cDNA Library Preparation: Life Technologies, Inc.
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLM11139 row: 0 column: 07
               High quality sequence stop: 849.
               Location/Qualifiers
                 1..853
                   /organism="Mus musculus"
                   /strain="FVB/N"
                   /db_xref="taxon:10090"

```

```

/clone="IMAGE:5051358"
/lab_host="NCI_CGAP_L19"
/note="Organ: Liver; Vector: PCMV-SPORT6; Site_1: NCI;
Site_2: Sall; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      243 a      199 c      194 g      217 t
ORIGIN

```

```

alignment_scores:
    Quality: 1015.50      Length:      287
    Ratio: 4.474          Gaps:      5
    Percent Similarity: 79.094      Percent Identity: 70.383

```

```

alignment_block:
US-09-980-881-2 x B1144644

```

```

Align seg 1/1 to: B1144644 from: 1 to: 853

```

```

3 LeucylSerLeuAlaValLeuValProIleValLeuPheCysGluGlnI 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 CTTCATGGCCTTGGAATCCTGTAGCCATCATCCGCTAT...GAGCAGCA 48
19 sValPheAlaPheGlnSerGlyGlnValAlaIleLeuProArgThr 36
| | | | | | | | | | | | | | | | | | | | | | | | | | |
49 TGCTTCGCTTCAGAGTGCCAGGCTTTATCTGCTCTCCAGAAACCT 98
36 eirYrGlnValGlnValLeuGlnAsnLeuThrPheTrpGluIleVal 52
| | | | | | | | | | | | | | | | | | | | | | | | | | |
99 CCAAGGCAAGTCACTACTAGATCTTACTACAGCTATGAGAGTGGT 148
53 LeuTrpGlnProValThrAlaAspLeuIleValIleCysGlnValAl 69
| | | | | | | | | | | | | | | | | | | | | | | | | | |
149 CTCTGGAGCAGCAGCAGCAGCAGTCAATCATGAGAGAAAGAAAGCA 198
69 sPhePheValAsnAlaSerAspValAspAsnValIleAlaHisLeu 86
| | | | | | | | | | | | | | | | | | | | | | | | | | |
199 CTTTTTGTAAATGCGTCTGATGCGACAGTGTCAAAAGCCGCTTAA 248
86 aIleGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 102
| | | | | | | | | | | | | | | | | | | | | | | | | | |
249 TGAGCAGAAATTCATTACGTTCTGATGACACAGTGGAGACCTAAT 298
103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAla 119
| | | | | | | | | | | | | | | | | | | | | | | | | | |
299 GAACAGACAGCTTTCATGACAGCGTCAGCCCGCGCTCCGCTCATA 348
119 rYrGluGlnTrpHisSerLeuAsnGluIleTrpSerTrpIleGlu 136
| | | | | | | | | | | | | | | | | | | | | | | | | | |
349 CTATGACACATATCATCTGCTAAATGAAATCTATCTGATAGATCA 398
136 lPheTrpGlnAlaGlnHisProAspMetLeuThrIleHisIleGly 152
| | | | | | | | | | | | | | | | | | | | | | | | | | |
399 TAACTGACACAGCATCTGACATGCTCCAGAAATCTACATCGATCA 448
153 PheGluLysTrpProLeuTrpValLeuLysValSerGlyGlnGln 169
| | | | | | | | | | | | | | | | | | | | | | | | | | |
449 TTGAGAGAAATGACCATTTATGTTTAAAGGTCTCAGAGAAAGCA 498
169 rAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgIle 186
| | | | | | | | | | | | | | | | | | | | | | | | | | |
499 AATCAAAATGCACTGTGATCGACTGTGAAATCCATGCCAGAAATG 548
186 lSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
| | | | | | | | | | | | | | | | | | | | | | | | | | |
549 TTTCACCTGCTTCTCTGTTGTGTTCAATAGGCTACGTGACAAAT 598
197 ..... 197
599 GGGAAGAAATCTGTATACAGACTTCTGAGCAGCGATTTCTACAT 648

```

```

198 .....Asn 199
649 CATGCCGCTGATGAACGTGATGCTATGACTACACGTGAAAAAGATC 698
199 rgMetTrpArgLysAsnArgSerPheTrpIleAsnAsnHisCysIle 215
| | | | | | | | | | | | | | | | | | | | | | | | | | |
699 GAATGTGAGAGAAACCGT...CTGTCAACAAGAACACCGTGGTGGG 745
216 ThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluGln 232
| | | | | | | | | | | | | | | | | | | | | | | | | | |
746 CCAGACCTGAACAGAACTTCCTCCAAACACTGTGTGAGAAAGTGC 795
232 aSerSerSerSerCysSerGluThrTrpCysGlyLeuTrpProGlu 248
| | | | | | | | | | | | | | | | | | | | | | | | | | |
796 GTCCAGTTCCTGCTGCTGAAACCTACTGTGGAGCTTATCTGATC 845
248 rGluPro 250
| | | | |
846 TGAGCCG 852

```

```

seq_name: gb_est2:BG616456

```

```

seq_documentation_block:

```

```

LOCUS      BG616456              738 bp      mRNA      linear      EST_18-APR-2001
DEFINITION 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',

```

```

ACCESSION  BG616456
VERSION    BG616456.1 GI:13667827

```

```

KEYWORDS  EST.
SOURCE     human.

```

```

ORGANISM  Homo sapiens

```

```

REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

TITLE     1 (bases 1 to 738)

```

```

JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)

```

```

COMMENT   Unpublished (1999)

```

```

Contact: Robert Strausberg, Ph.D.

```

```

Email: cgabbs-r@mail.nih.gov

```

```

tissue Procurement: CLONETECH Laboratories, Inc.

```

```

cDNA library Arrayed by: rpe I.M.A.G.E. Consortium (LLNL)

```

```

DNA Sequencing by: Incyte Genomics, Inc.

```

```

Clone distribution: MGC clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/LLNL at:

```

```

http://image.llnl.gov

```

```

Plate: LLCMI595 row: c column: 05

```

```

High quality sequence stop: 666.

```

```

FEATURES

```

```

source

```

```

1..738

```

```

/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"

```

```

/clone="IMAGE:4733500"

```

```

/lab_host="NCI_MGC_76"

```

```

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:

```

```

SfiI (ggcgctcgcc); Site_2: SfiI (ggccatagcc); 5' and

```

```

3' adaptors were used in cloning as follows: 5' adaptor

```

```

sequence: 5'-CACGGCATTTATGGCC-3' and 3' adaptor sequence:

```

```

5'-ATTCTAGAGCCGACGCGCCGACATG-dt(30)BA-3' (where B = A,

```

```

C, or G and N = A, C, G, or T). Average insert size 1.85

```

```

kb (range 1.0-4.0 kb). 15/15 colonies contained inserts

```

```

by PCR. This library was enriched for full-length clones

```

```

and was constructed by Clontech Laboratories (Palo Alto,

```

```

CA). Note: this is a NIH_MGC Library."

```

```

BASE COUNT      209 a      161 c      153 g      215 t
ORIGIN

```

```

alignment_scores:
    Quality: 996.50      Length:      296
    Ratio: 4.449          Gaps:      4
    Percent Similarity: 75.676      Percent Identity: 69.257

```


Fax: 86-21-50801922
Email: hanzge@nc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers

FEATURES

1..735

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCMTD06"
/clone_11b="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 200 a 158 c 172 g 205 t
ORIGIN

alignment_scores:
Quality: 949.00 Length: 193
Ratio: 5.048 Gaps: 1
Percent Similarity: 97.409 Percent Identity: 96.891

alignment_block:
US-09-980-881-2 x AV646979 ..

Align seg 1/1 to: AV646979 from: 1 to: 735

```

6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAl 22
|||||
7 CTTCGCGCTCTGTACCCATGTTCTCTCTGTGACAGACATGCTCTCCG 56
|||||
22 apheGlnSerGlyGlnValLeuAlaLeuProAArgThrSerArgGln 39
|||||
57 GTTTCAGAGTGGCCAAAGTCTACGCTCTCTCTACAGCCTTAGGCAAG 106
|||||
39 aIGlnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTyrGln 55
|||||
107 TTCAGTCTTACAGATCTTACACAAATATGAGATTGCTCTCTGCGAG 156
|||||
56 ProValThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 72
|||||
157 CCGGTAACGGTGACCTTATGTGAAGAAAAAACAAGTCCATTTTTCGT 206
|||||
72 IAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly 89
|||||
207 AATGCAATGATGATGTCGACAAATGTGAAGCCCAATTAATGTGAGCGGA 256
|||||
89 IeProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 105
|||||
257 TTCATGCGAGTGTCTGTGGCAGAGCTGGAAGATCTTATTCACAGCAG 306
|||||
106 ILeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 122
|||||
307 ATTTCACAGACAGACAGTCCAGCCCGAGGCTCCGCATCGCATATGACA 356
|||||
122 nTyrHisSerLeuAsnGluIleTyrSerTyrPheIleupheIleThrGlu 139
|||||
357 GATACACTACATAAATGAATCTATCTTGGATAGAAATTTAATACCTGAGA 406
|||||
139 rGHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 155
|||||
407 GGCATCTCGATGCTTACAAAATCCACATGATGATCCATTTGAGAAAG 456
|||||
156 TyrProLeuTyrValLeuLys ValSerGlyLysGluGlnThrAlaLysA 172
|||||
457 TACCCACTCTATGTTTAAAGGGTGTGGAAAGAACAAACAGCCCAAA 506
|||||
172 snAlaIleTyrPheAspCysGlyIleHisAlaArgGluTyrPheSerPro 188
|||||
507 ATGCCATATGATGATGCTGTGAATCCATGCCAGAAATGATCTCTCTCT 556
|||||
189 AlaPheCysLeuTyrPheIleGlyHis 197

```

||||| ||| |||||
557 GCTTCTGTGGGGGTCATAGGCGCAT 583

seq_name: gb_est2:BG618813

seq_documentation_block:

LOCUS BG618813 750 bp mRNA linear EST 18-APR-2001
DEFINITION 602646186F1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4767885 5',
MRNA sequence.
ACCESSION BG618813
VERSION BG618813.1 GI:13670184
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 750)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LUCM1630 row: k column: 22
High quality sequence stop: 622.

FEATURES

source

Location/Qualifiers
1..750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4767885"
/clone_11b="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGACC-3' and 3' adaptor
sequence: 5'-ATTCTAAGGCCGAGCGCCGAGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

BASE COUNT 219 a 146 c 171 g 213 t
ORIGIN

alignment_scores:
Quality: 927.50 Length: 210
Ratio: 4.781 Gaps: 5
Percent Similarity: 92.381 Percent Identity: 90.476

alignment_block:
US-09-980-881-2 x BG618813 ..

Align seg 1/1 to: BG618813 from: 1 to: 750

```

1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysG 17
|||||
11 ATGAAGCTTTCAGAGCTTCGAGTCTCTGTACCCATGTTCTCTCTGGA 60
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuProA 34
|||||
61 GCAGCATCTCTTCGGGTTTCAGAGTGGCCAAAGTCTACGCTCTTCCTA 110
|||||
34 rGThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyrGlu 50
|||||
111 GAACCTTAGGCAAGTTCAGTCTTACAGATCTTACACAAATATAGAG 160

```

```

51 11leValleuTrpGlnProValThrAlaAspLeuIleVallylsLysG1 67
|||||
161 ATGCTCTCTGGCAGCGGTACAGCTGACCTTATGTGTGAAGAAAAACA 210
67 nvalHisPhePheValAlaAsnAspValAspVallylsLysL 84
|||||
211 AGTCCATTTTGTGTAAATGATCTGATGTCGACAAATGTGAAGCCATT 260
84 euAsnValSerGlyIleProCysSerValIleuAlaAspValGluAsp 100
|||||
261 TAATGTGACGCGAATCCATGCGAGTGTCTTCTGCGCAGACGTGAGAT 310
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
311 CTATATCAACAGCAGATTTCCACAGACAGACAGACCCCGAGCCCGC 360
117 aSerTyTrpGlnGlnIleThrHisSerLeuAsnGlnIleTySerTrpIleG 134
|||||
361 ATGCTACTATGACAGATCATCTCCTAATGAATCATCTATCTTGATAG 410
134 luePheIleThrGluArgHisProAspMet.LeuThrIleHisIleG1 150
|||||
411 AATTATTAAGTGAAGGATCCCTATATGCTTAACAAAAATCCACATTGG 460
150 y.SerSerPheGluLysTyTrpLeu.TyrValIleuLysValSer.GlyL 166
|||||
461 CATCTCATTCGACAGACATCAACAGTCGATGTTTAAAGTTCTCGGAA 510
166 ysgGlnIleThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
|||||
511 AAGAACAGACAGCGCAAAATGCCATATGATGATGACTGTGAAATCCATGCC 560
183 ArgGluTrpIleSerProAlaPheCysLeuTrp..... 193
|||||
561 AGAATAAGATCTCTCTGCTTCTGCTGTGTGTCATAGGCCATATACT 610
194 ....PheIleGlyHisAsnArg 199
|||||
611 CAATTTCTATGCGATATAG 632
seq_name: gb_estc2:BG565100

seq_documentation_block:
LOCUS BG565100 747 bp mRNA linear EST 10-APR-2001
DEFINITION 602583720F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711348 5',
mRNA sequence.
ACCESSION BG565100
VERSION BG565100.1 GI:13572753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 747)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CM151 row: h column: 05
High quality sequence stop: 714.
Location/Qualifiers
1..747
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="IMAGE:4711348"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1:
5'f1 (ggcgctggcc); Site: 2: 5'f1 (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATATGACC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCGGCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 220 a 154 c 166 g 207 t
ORIGIN

alignment_scores:
Quality: 902.00 Length: 251
Ratio: 4.422 Gaps: 6
Percent Similarity: 81.275 Percent Identity: 77.291

alignment_block:
us-09-980-881-2 x BG565100 ..

Align seg 1/1 to: BG565100 from: 1 to: 747

52 ValLeuTrpGlnProValThrAlaAspLeuIleVallylsLysGlnVa 68
1 GTTCTCTGCGACCGCGTACAGCTGACCTTAT.GTGAAGAAAAACAAGT 49
68 lHisPhePheValAlaAsnAspValAspVallylsLysLysLeuA 85
|||||
50 CCATTTTGTGAGATGATCGATGTCGACATGTGAAGCCATTAA 99
85 snValSerGlyIleProCysSerValIleuAlaAspValGluAspLeu 101
|||||
100 ATGTGAGCGGAATCCATGACAGTGTCTGCGCAGATGTGAAGATCTT 149
102 lIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAla 118
|||||
150 ATTCACACGCGATTTCCACAGCACAGTCAGCCCCCGAGCCCTCGCATC 199
118 rTyTrpGlnGlnIleThrHisSerLeuAsnGlnIleTySerTrpIleGln 135
|||||
200 GTACATGACAGATATCACTCATTAATGAATCATCTTGATAGAGAT 249
135 lleIleThrGluArgHisProAspMetLeuThrIleHisIleGlySer 151
|||||
250 TTATTAAGTGAAGGATCTCTATATGCTTAACAAAAATCCACATTGATCC 299
152 SerPheGluLysTyTrpLeuTrpValIleuLysValSerGlyLysGlnG 168
|||||
300 TCATTTGAGAGATACCACTCATGTTTAAAGTTTGTGAAGAAACA 349
168 nThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisLysArgGlu 185
|||||
350 AGCAGCCCAAAATGCCATATGATGATGACTGTGAGATTCATGCCAGAGAT 399
185 rPheIleSerProAlaPheCysLeuTrpPheIleGlyHis..... 197
|||||
400 GGACCTCTCTGCTTCTGCTGTGTGTCATAGGCGCATATCACTCATTC 449
197 ..... 197
450 TATGGATTAATAGGCAATATACATCTCTGAGGCTTGATGATTCTA 499
198 .....AsnArgMetTrpArg.... 202
500 TGTATGCCAGTGTATATGATGATGTTATGACTACTCATCAAGAAACAAG 549
203 .....LysAsnArgSerPheTyAlaAsnAsnHisGys 213
|||||

```

550 AATCGAATGGAGAACAGACCGTCTTCTATGCGAACATCATTC 599
 214 TlleclyThraspleuasensaraphelValSerlyshstPcysglu1 230
 600 ATCGAAGACACCTGAAATAGCAACTT.GCTTCCAAACACGCTGTGAGGA 648
 230 u.GlyAlasSerSerSer.CysSerGlu.Thltyr.CysGlyLeuTy 245
 649 AGGCTGATCCAGTCTCTCATTCGGAACCACTACTGTGACCTTAT 698
 246 ProgluSerGluProgluValLysAlaValAlaSerPheLeuArg 260
 699 CCGAGTCAGAACACGAGTG.GAAGCAGTGGCTAGTTCTTGAAG 742
 seq_name: gb_est2.BG568240

seq_documentation_block:
 LOCUS BG568240 691 bp mRNA linear EST 10-APR-2001
 DEFINITION 602587149F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715797 5',
 mRNA sequence.

ACCESSION BG568240
 VERSION BG568240.1 GI:13575893
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 691)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgepds-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
 http://image.lnl.gov
 Plate: L1CM1563 row: a column: 14
 High quality sequence stop: 691.
 Location/Qualifiers
 1. 691

FEATURES
 source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4715797"
 /clone_id="NIH_MGC_76"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggccgcctggcc); Site_2: SfiI (ggccatattggc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGCGCGCGCGCATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."

BASE COUNT 204 a 143 c 144 g 200 t
 ORIGIN

alignment_scores:
 Quality: 898.00 Length: 232
 Ratio: 4.582 Gaps: 4
 Percent Similarity: 84.483 Percent Identity: 81.034

alignment_block:
 US-09-980-881-2 x BG568240 ..
 Align seg 1/1 to: BG568240 from: 1 to: 691

1 MetlylsLeucYsserLeuAlaValleuValProIlleValleuPhecysgl 17
 19 ATGAGCGTTTGACAGCCCTTGACAGCTGCTGTACCCATGTTCTCTCTGTA 68
 17 uGlnHisValPheAlaPheGlnSerGlyGlnValleuAlaAlaLeuPro 34
 69 GCAGCATGCTTCCGCGTTTCAGAGTGGCCAACTTACGCTCTTCCCA 118
 34 rGthSerArgGlnValGlnValleuGlnAsnLeuThrThrTrpTrpGlu 50
 119 GAACCTTACGCAAGTTCAGATTCTACAGAACTTCTACACATATGAG 168
 51 TleValleuTrpGlnProValThrAlaAspLeuIleValLysLysGsl 67
 169 ATGTCTCTGCGCAGCGGTAAACAGCTGACCTTAT.GTGAGAGAAAAACA 217
 67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
 218 AGTCACATTTTGTAAATGACATGATGTCGACAAATGTGMAAGCCCAT 267
 84 euAsnValSerGlyIleProCysSerValleuLeuAlaAspValGluasp 100
 268 TAAATGTAGCGGAATTCATGACATGCTCTGTCGACAGCGTGAAGAT 317
 101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
 318 CTATATCAACAGCAGATTCTCAACGACACAGCAGCCCGGACCTCCGC 367
 117 aserTrpTrpGlnGlnTrpHisSerLeuAsnGlnIleTrpSerTrpIleG 134
 368 ATGCTACTATGAAACAGTATACCTCACTAAATGAAATCTTCTTGATG 417
 134 lPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
 418 AATTATTAACAGAGAGCATTCGATATCTTACAAATTCACATGGA 467
 151 SerSerPheGlnLysTrpProLeuTrpValleuLysValSerGlyLysGsl 167
 468 TCCTCATTTAGAGAAATGACACCTCATGTTTAAAGGTTTCGAAAAAGA 517
 167 uGlnThrAlaLysAsnAlaIle..... 174
 518 ACAAGACGCCAAATATGCCATTAACCTAATTCATGAGATTAAGGCG 567
 175TripleaspCysGlyIleHisAlaArgGlu 184
 568 AATATACCAATCTCCGAGCGTTGTGATTC..... 599
 185 TripleSerProAlaPheCysLeu.....TrpPheI 195
 600 TATGTTATGCCGCGTGAATGATGATGGTATGACTACTCATG..... 644
 195 eGlyHisAsnArgMetTrpArgLysAsnArgSerPheTrpAlaAsn 210
 645 .AAACGAATCGAATGTGAGAAACACCGCTTCTTATGCGAAG 689
 seq_name: gb_est1:A1048167

seq_documentation_block:
 LOCUS A1048167 724 bp mRNA linear EST 08-JUL-1998
 DEFINITION ud71b12.y1 Sugano mouse liver m1a Mus musculus cDNA clone
 IMAGE:1451327 5' similar to TR:015114 Q15114 PREPRO-PLASMA
 CARBOXYPEPTIDASE B. / mRNA sequence.

ACCESSION A1048167
 VERSION A1048167.1 GI:3296454
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 1 (bases 1 to 724)
 Maira,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project

FEATURES	Location/Qualifiers
source	1. .724

BASE COUNT ORIGIN	215 a	169 c	160 g	178 t	2 others
----------------------	-------	-------	-------	-------	----------

```
alignment_scores:
  Quality: 884.50      Length: 241
  Ratio: 4.631        Gaps: 1
Percent Similarity: 79.253  Percent Identity: 70.124
```

```
alignment_block:
US-09-980-881-2 x AI048167 .
```

Align seg 1/1 to: AI048167 from: 1 to: 724

23 pHeGInserGlyGlnValLeuAlaAlaLeuProKgrHrSsrArgGlnVa 39
2 TTTCAGAGGGCCAGGCTTTTATCTGCTCTTCCAGAACTCCAGCAGT 51
39 IclnValLeuGlnAsnLeuThrThrThrTyrgluIleValLeuTrpGlnP 56
52 TCACCTACTCTCGAATCTTACTACAACTTAAGAGCTGGTCTCGCAGC 101
56 rGValThrAlaAspLeuIleValLysLysLysGlnValLHisPhePheVal 72
102 CAGTGCACACTGAATCTTCGCGAAGAAAGAAAGGAAGTCCACTTTTGTG 151
73 AsnAlaSerAspValAspAsnValLysLysLHisLeuAsnValSerGlyI 89
152 AATGCCTCTGAGTGTGCACAGGTCCAAAGCCGCAATTAAATGTGACACAGT 201
89 eProCysSerValLeuLeuAlaAspValGlnAspLeuIleGlnGlnI 106

202	TCGATTAAACGTTCTGATGAMACAGCGTGGAGACCTAATTGACAGCA	251
106	leSerAsnAspThrValSerProAlaGlyAlaSerAlaSerTyrTyrGluGln	1222
252	CTTTCATGACACAGGTGACCCCGCCGCGCTTCCTACTATGAGCAG	3010
123	TyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluAr	139
302	TATCACCCTCGTAAATGAAATCTATTCTCGTGGATAGAAAGCATAACTGACA	3515
139	GHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluTyrSt	156
352	GCATCCGACATCTCCAGAAATATCACTGCATCATTCGAGAAAGT	4010
156	YrProLeuTyrValLeuLysValSerGlyGluGlnThrAlaLysAsn	1722
402	ACCCACTTATGTTTAAAGGCTCAGGACAGGAACAAAGATCAAAAT	4515
173	AlaIleTrpIleAspCysGlyIleHisAlaIleArgIleTrpIleSerProAl	189
452	GCCATCGGATGTCAGCTGTGGATCCATCGCAGAGATGATTCACCTGC	5010
189	ApheCysLeuTrpPheIleGlyHis	197
502	TCCTGTGTTGTGGTTCATGAGCTACGTGACACAAATTCATGGAGAGAA	5515
197	197
552	ATCTGATACACAGCTTTCGAGGACGCTGGATCTCTCATCANTGACAGTG	6010
198	198
602	ATGACAGCTGGATGGCATATACATACACCGGAAAAAGATTCAAATGTGGAG	6515
202	GlyAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuA	219
652	GAGAGCGGCTGTGCTCACACAAACACACGCGTGGGACACAGCTGA	7010
219	snSerAsnPheValSerLysHis	226
702	ACAGGAACCTTCGCTTCCAACAC	724

seq_name: gb_est1:AV693037

seq_documentation_block:

LOCUS	AV693037	583 bp	mRNA	linear	EST 16-JAN-2002
DEFINITION	AV693037	GKC Homo sapiens cDNA clone GKCAPH12.5', mRNA sequence.			
DESCRIPTION	AV693037				

VERSION AV6993037.1 GI:10294900
KEYWORDS EST.
SOURCE 1

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

FEATURES

SOURCE

```
1. .583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCAFH12"
/clone_1kb="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
```

BASE COUNT 161 a 130 c 116 g 176 t
ORIGIN

alignment_scores:

Quality: 868.00 Length: 197
Ratio: 4.932 Gaps: 1
Percent Similarity: 89.340 Percent Identity: 87.817

alignment_block:

US-09-980-881-2 x AV693037 ..

Align seg 1/1 to: AV693037 from: 1 to: 583

```
1 MetIysLeuCySerLeuAlaValLeuValProIleValLeuPheCysG1 17
|||||
16 ATGAAGCTTTCAGAGCTTGCAGCTCTGTACCATGTTCTCTCTGTGA 65
|||||
17 uGlnHisValPheAlaPheGlnSerGlyGlnValLeuAlaLeuPro 34
|||||
66 GCAGCATGCTTCGCCCTTCAGAGTCCAGTCCAGTTCAGCTCTTCCTA 115
|||||
34 rGhrSerArgGlnValGlnValLeuGlnAsnLeuThrThyTyrGlu 50
|||||
116 GAACCTCTAGGCAAGTTCAAGTCTACAGATCTTACTACAAACATATGAG 165
|||||
51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysG1 67
|||||
166 ATGTGCTCTGCGACGCCGGGTACAGCTTATGTGAGAAAAACA 215
|||||
67 nValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisL 84
|||||
216 AGTCCATTTTGTGTAATGCATCTGATGCGCAATGTGAAGCCCATTT 265
|||||
84 euAsnValSerGlyIleProCySerSerValLeuLeuAlaAspValGluAsp 100
|||||
266 TAAATGTGAGCGGAATTCATGCGATGCTGCTGCGACAGCTGGAAGAT 315
|||||
101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAl 117
|||||
316 CTATATTCACACAGCAGATTTCACACAGACAGTCAGCCCGAGCCCTCCGC 365
|||||
117 aSerTyrTyrGlnGlnTyrHisSerLeuAsnGlnIleTyrSerTrpIleG 134
|||||
366 ATCGTACTATGAACAGATACACTACATAATGAATCTATCTTGGATAG 415
|||||
134 LuPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
|||||
416 AATTATTAAGTACAGCATCTGATATGCTTACAAAAATCCACATTTGGA 465
|||||
151 SerSerPheGlnLysTyrProLeuTyrValLeuLysValSerGlyLysG1 167
|||||
466 TCCTCATTTTGAAGATACCCACTATGTTTAAAGGTTTCGGGAACA 515
|||||
167 uGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 184
|||||
516 GCC..... 518
184 LuTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
:::|||||
519 .....ACTTCTGCTGTGTGTTACATAGGCCAT 545
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: September 18, 2002, 16:47:48 ; Search time 23.83 Seconds

(without alignments)
549.190 Million cell updates/sec

Title: us-09-980-881-3

Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQL.....IKYFTSNPPEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	691.5	38.5	415	1 CBPB_RAT	P19223 rattus norv
2	637.5	35.5	417	1 CBPB_HUMAN	P19086 homo sapien
3	635.5	35.3	401	1 CBPB_PIG	P09955 sus scrofa
4	632	35.2	417	1 CBPC_MOUSE	P15089 mus muscula
5	630.5	35.1	416	1 CBPC_CANFA	P15261 canis famill
6	622	34.6	417	1 CBPC_HUMAN	P15088 homo sapien
7	586	32.6	306	1 CBPB_BOVIN	P00732 bos taurus
8	552	30.7	309	1 CBPC_RAT	P1961 rattus norv
9	549.5	30.6	417	1 CBPB_RAT	P19222 rattus norv
10	536.5	29.8	419	1 CBPA_BOVIN	P00730 bos taurus
11	533.5	29.7	419	1 CBP1_RAT	P00731 rattus norv
12	511.5	28.4	417	1 CBP2_HUMAN	P48052 homo sapien
13	493.5	27.4	419	1 CBP1_HUMAN	P15085 homo sapien
14	491	27.3	421	1 CBP4_HUMAN	P42788 simullum vi
15	411.5	22.9	304	1 CBPA_MOUSE	P04069 anophelus g
16	393	21.9	433	1 CBPA_ANOGA	P04069 anophelus g
17	375.5	20.9	303	1 CBPB_ASTEL	P28068 saccharomyc
18	300	16.7	430	1 YHT2_YEAST	P28068 thermactin
19	286.5	15.9	424	1 CBPT_THEVU	P16143 streptomyc
20	283	15.7	451	1 CBPS_STRGR	P39041 streptomyc
21	259	14.4	434	1 CBPS_STRCP	P54789 schizosacch
22	116.5	6.5	707	1 ORC1_SCHPO	P70080 gallus gall
23	103.5	5.8	445	1 TRSH_CHICK	P79005 schizosacch
24	101	5.6	663	1 TAZ1_SCHPO	P54497 bacillus su
25	99	5.5	376	1 YOGT_BAGSU	P17552 homo sapien
26	99	5.5	444	1 TRSH_HUMAN	O15055 homo sapien
27	99	5.5	1255	1 PER2_HUMAN	O15006 homo sapien
28	97	5.4	297	1 Y103_HUMAN	P17290 oryctolagus
29	95	5.3	444	1 TRSH_RABIT	O45706 bacillus th
30	94	5.2	1160	1 CBPA_BACTP	P27206 bacillus su
31	93.5	5.2	3587	1 SRP1_BACST	P23615 saccharomyc
32	93	5.2	1451	1 SRP6_YEAST	P37892 lophius ame
33	92	5.1	454	1 CBPH_LOPAM	

34	92	5.1	488	1	TY3H_ANGAN	O42091 anguilla an
35	91	5.1	262	1	YCU1_ECOLI	P51983 escherichia
36	90.5	5.0	319	1	UN30_CAEEL	P52906 caenorhabdi
37	90.5	5.0	434	1	CBPB_BOVIN	P04836 bos taurus
38	90.5	5.0	476	1	CBPB_RAT	P15087 rattus norv
39	89.5	5.0	476	1	CBPB_MOUSE	O00493 mus musculi
40	89	4.9	481	1	TRSH_XENLA	O92142 xenopus lae
41	89	4.9	575	1	ACBEA_YEAST	O12031 saccharomyc
42	88.5	4.9	476	1	CBPB_HUMAN	P16870 homo sapien
43	88	4.9	447	1	TRSH_MOUSE	P17532 mus musculi
44	88	4.9	523	1	SYK_PYRHO	O57963 pyrococcus
45	87.5	4.9	396	1	ENP1_BACSH	O03415 bacillus sp

ALIGNMENTS

RESULT	ID	CBPB_RAT	STANDARD:	PRT:	415 AA.
AC	P19223;				
DR	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Carboxypeptidase B precursor (EC 3.4.17.2).				
GN	CPB.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=89034324; PubMed=3182872;				
RX	Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;				
RT	Structural characterization of the rat carboxypeptidase A1 and B				
RT	genes. Comparative analysis of the rat carboxypeptidase gene				
RT	family.";				
RL	J. Biol. Chem. 265:17837-17845(1988).				
CC	-I- CARBOXYPEPTIDASE ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =				
CC	peptide + L-lysine(or L-arginine).				
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE				
CC	ZINC CARBOXYPEPTIDASE FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/				
CC	or send an email to license@sdb.ch).				
CC	-----				
DR	EMBL; M23959; AAA40872.1; JOINED.				
DR	EMBL; M23947; AAA40872.1; JOINED.				
DR	EMBL; M23950; AAA40872.1; JOINED.				
DR	EMBL; M23952; AAA40872.1; JOINED.				
DR	EMBL; M23953; AAA40872.1; JOINED.				
DR	EMBL; M23954; AAA40872.1; JOINED.				
DR	PIR; A3129; A32129.				
DR	HSSP; P09955; INSA.				
DR	MEROPS; M14.003; --.				
DR	InterPro; IPR003146; Proper_M14.				
DR	InterPro; IPR000834; Zn_carboxpept.				
DR	Pfam; PF02244; Proper_M14; 1.				
DR	Pfam; PF00246; Zn_carboxpept. 1.				
DR	PRINTS; PR00765; CRBOXYPLASEA.				
DR	PROSITE; PS00132; CARBOXYPEPT_2N.1; 1.				
DR	PROSITE; PS00133; CARBOXYPEPT_2N.2; 1.				
KW	Hydrolase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.				
FT	SIGNAL	1	13		
FT	ACTIVATION PEPTIDE.	14	108		
FT	CHAIN	109	415		
FT	FT	174	174		
FT	METAL	177	177		
FT	METAL	177	177		
FT	ZINC (BY SIMILARITY).				
FT	ZINC (BY SIMILARITY).				

FT METAL 302 302 ZINC (BY SIMILARITY).
 FT ACT_SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 171 184 BY SIMILARITY.
 FT DISULFID 243 266 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 SQ SEQUENCE 415 AA; 47515 MW; 8EA06CCADE30B6F2 CRC64;

Query Match 38.5%; Score 691.5; DB 1; Length 415;
 Best Local Similarity 40.2%; Pred. No. 8.7e-52;
 Matches 145; Conservative 59; Mismatches 114; Indels 43; Gaps 6;

OY 1 FOSGOVLALPRTSRQVOYLONTTTEIYLMOPVTADLYKKQVHFVNASDVNDKA 60
 DB 20 FDGNVYVSVHGEDHVNLDIELANTKEIDFWKPSAQVPLTTVDHFVAEDVADVEN 79
 OY 61 HLNVSIGPVSLLADVEDLIDOOISNDVSPRASAYEYHSLNEIYSWIEFITERHPD 120
 DB 80 FLEENEVHYEVLISVNRNALSQFDSHT--RASGHSTYKTKNKMETIEMWIOQVATNDPD 136
 OY 121 MLTKHIGSFEEKYPLYLVKSGKEQTANNAIWIDCGIHAREWISPAFCMLFT----- 173
 DB 137 LVTSVIGTTEGGRNMYLKI-GKTRPNKPAIFIDCGFHAREWISPAFCQMFVREAVRTY 195
 OY 174 -----NRMWRKRSFYANNHCIGTDLNRNFSK 203
 DB 196 NOEIHMQLDLDELDFYLPVYNIDQIYTTWTRDKMRKTRSTMGSSCLGVPRPNRF-NA 254
 OY 204 HWCSEGAASSSCSEYTCGLYPESEPEYKAVASFLRNINQIKAYISMYSQHIYFYSY 263
 DB 255 GWCCEGASRSPSEYTCGPAPSESEKELADIFRNINLSTIKAVLTHISYQMLLYPSY 314
 OY 264 TRSKSHOHEELSLVASEVRAIEKTSKTRTHGHGSETLLAAGGDDWYIDGIXYSF 323
 DB 315 DYKLEPTEELNALVAKAKEL-ATLHGQTYTYGPGATTIYPAAGSDWMSYDQIKYSF 373
 OY 324 T 324
 DB 374 T 374
 RESULT 2
 CBPB_HUMAN STANDARD; PRT; 417 AA.
 ID CBPB_HUMAN
 AC P150B6; 060834;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein) (PASP).
 GN CPB1 OR CPB OR CPBP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=92129345; PubMed=1370825;
 RA Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S., French C.K.;
 RT Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B.";
 RL J. Biol. Chem. 267:2575-2581(1992).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=98182241; PubMed=9524066;
 RA Aloy P., Catasus L., Villegas V., Reverter D., Vendrell J., Aviles F.X.;
 RT Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B.";

RL Biol. Chem. 379:149-155(1998).
 RN [3]
 RP SEQUENCE OF 16-43.
 RC TISSUE=Pancreas;
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT Purification and properties of five different forms of human procarboxypeptidases.";
 RL Eur. J. Biochem. 179:609-616(1989).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2O) = peptide + L-lysine(or L-arginine).
 CC -1- TISSUE SPECIFICITY: PANCREAS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE ZINC CARBOXYPEPTIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; M81057; AAA66973.1; .
 DR EMBL; AJ224866; CAAL2163.1; .
 DR PIR; S02812; S02812.
 DR PIR; A42332; A42332.
 DR HSPB; P09955; INSA.
 DR MEROPS; M14.003; .
 DR MIM; 114852; .
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; zn_Carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; zn_Carboxypept; 1.
 DR PRINTS; PR00765; CARBOXYPTASEA.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 DR KMW; Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 110
 FT CHAIN 111 417
 FT METAL 176 176
 FT METAL 179 179
 FT METAL 304 304
 FT ACT_SITE 378 378
 FT DISULFID 173 186
 FT DISULFID 245 268
 FT DISULFID 259 273
 FT CONFLICT 16 16
 FT CONFLICT 17 17
 FT CONFLICT 37 37
 FT CONFLICT 208 208
 FT CONFLICT 245 245
 SQ SEQUENCE 417 AA; 47366 MW; BBICF212D830305E CRC64;

Query Match 35.5%; Score 637.5; DB 1; Length 417;
 Best Local Similarity 36.0%; Pred. No. 3.7e-47;
 Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

OY 1 FOSGOVLALPRTSRQVOYLONTTTEIYLMOPVTADLYKKQVHFVNASDVNDKA 60
 DB 22 FEGEKYFRVAVEDENHINIRELASTQIDFWKPSAQVPLTTVDHFVAKEDVTYVEN 81
 OY 61 HLNVSIGPVSLLADVEDLIDOOISNDVSPRASAYEYHSLNEIYSWIEFITERHPD 120
 DB 82 VLKQNELOKRYVLISNRYNVEAQFDSRV--RATGHSYKTKNKMETIEMWIOQVATNDPA 138
 OY 121 MLTKHIGSFEEKYPLYLVKSGKEQTANNAIWIDCGIHAREWISPAFCMLFT----- 173
 DB 139 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQMFVREAVRTY 197
 OY 174 -----GH-----NRMWRKRSFYANNHCIGTDLNRNFSK 203

```

Db 198 GREIOVTELLKLDYFVLPVINDGIYTWKSRFRWTRSRHTGSSCGTDPNRNF-DA 256
QY 204 HMCCEGASSSSCSEYTCGLYPESEPEVKAVASFLRNINQIKAYISMHSYQHIPEPSY 263
    ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 GWCETGASRNCDDEYTCGPAASEKETKALDFIRNNLSIKAYLTIHYSQMIIPYSY 316
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWIIDLGIKYSF 323
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 317 AYKLENNNAELNALAKATVKEL-ASLHGTGYKYYGPAATYIPAAAGSDMAYDQIGIRYSF 375
QY 324 T 324
Db 376 T 376

RESULT 3
CBPB_PIG STANDARD: PRT: 401 AA.
AC P09935;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID:9623;
RN [1]
RP SEQUENCE OF 1-103.
RX MEDLINE=91208150; PubMed=2018774;
RA Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA Aviles F.X.;
RT "Analysis of the activation process of porcine procarboxypeptidase B
RL Biochemistry 30:4082-4089(1991).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-38.
RX MEDLINE=85279427; PubMed=4026847;
RA Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
RT "Sequential homologies between procarboxypeptidases A and B from
RL porcine pancreas.";
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=91114690; PubMed=1989878;
RA Coll M., Guasch A., Aviles F.X., Huber R.;
RT "Three-dimensional structure of porcine procarboxypeptidase B: a
RL structural basis of its inactivity.";
RN [4]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX MEDLINE=91027767; PubMed=2223783;
RA Vendrell J., Wider G., Aviles F.X., Wuehrich K.;
RT "Sequence-specific 1H NMR assignments and determination of the
RL pancreatic structure for the activation domain isolated from
RL Biochemistry 29:7515-7522(1990).
RN [5]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX MEDLINE=91114693; PubMed=1989879;
RA Vendrell J., Billeter M., Wider G., Aviles F.X., Wuehrich K.;
RT "The NMR structure of the activation domain isolated from porcine
RL procarboxypeptidase B.";
RN [6]
RP EMBO J. 10:11-15(1991).
RN [7]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX MEDLINE=93044373; PubMed=1422143;
RA Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
RA Huber R., Wuehrich K.;
RT "Comparison of the NMR solution structure with the X-ray crystal
RL structure of the activation domain from procarboxypeptidase B.";

```

```

RL J. Biol. NMR 2:1-10(1992).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC Peptide + L-Lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -1- DATABASE: NAME-Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR PIR; B29181; B29181.
DR PDB; 1PBA; 31-OCT-93.
DR PDB; 1NSA; 24-DEC-97.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_CarDopept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carDopept; 1.
DR PRINTS; PR00765; CARBOXYPTASE.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR KMW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KM 3D-structure.
FT PROPEP 1 95 ACTIVATION PEPTIDE.
FT CHAIN 96 401 CARBOXYPEPTIDASE B.
FT DISULFID 158 171
FT DISULFID 230 253
FT DISULFID 244 258
FT METAL 161 161 ZINC.
FT METAL 164 164 ZINC.
FT METAL 289 289 ZINC.
FT ACT_SITE 341 341
FT ACT_SITE 363 363 NUCLEOPHILE.
FT STRAND 14 17
FT HELIX 20 31
FT TURN 32 32
FT HELIX 43 45
FT STRAND 50 52
FT HELIX 61 69
FT TURN 70 72
FT STRAND 75 76
SQ SEQUENCE 401 AA; 45713 MW; 53129AF159A26348 CRC64;

Query Match 35.3%; Score 635.5; DB 1; Length 401;
Best Local Similarity 37.4%; Pred. No. 5,2e-47;
Matches 135; Conservative 64; Mismatches 119; Indels 43; Gaps 6;

OY 1 FOSGOVIALALPRISROVOYIOMITTYEIVLQOPVADLIYKKOVHFPVNASDVONVKA 60
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 7 FEGEKYFRVNWEDENDISELHELASTROIDFKPDSVYQIKRSHYDFYKADILAVED 66
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 61 HNVSGIPCSVLADVEDLIQOQISNDTVSPRASASYEYHSLNIEYSMIEFIERHPD 120
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 67 FLEONELQYEVLIINLRSLVLAQFDSRC---RTHGSHYKYNWMTIEAMTEQVTSKND 123
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 121 MTKIHIGSFEKYPYLYVKSGKEQTAKNAIMIDCGIARREISAFCLMFI----- 173
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 LLSRSAIGTTFGDNITLYLKV-GKPSGNPKPAIFMDCGFHAREISQAFQWYRDAVRTY 182
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 174 -----GH-----NRMKRSFYANNHCIGTDLNRNFASK 203
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 183 GYEAHTEFLNDIYFVLPVINDGIYTWKSRFRWTRSRHTGSSCGTDPNRNF-NA 241
OY 204 HMCCEGASSSSCSEYTCGLYPESEPEVKAVASFLRNINQIKAYISMHSYQHIPEPSY 263
    ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 GWCETGASVNPCEYTCGSAASEKETKALDFIRNNLSIKAYLTIHYSQMIIPYSY 301
OY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSEETLYLAPGGDDWIIDLGIKYSF 323
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 DYKLPENDAELNSLAGAVKEL-ASLYGTYSYSGPSTIYIPAAAGSDMAYNOGIRYSF 360
OY 324 T 324
Db 361 T 361

```

```

RESULT 4
CBPC_MOUSE STANDARD: PRT: 417 AA.
ID CBPC_MOUSE
AC P15089;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Serafini W.E.;
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT novel member of the carboxypeptidase gene family."
RL J. Biol. Chem. 264:20094-20099(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05118; AAA37369.1; -
DR PIR: A34487; A34487.
DR HSSP: P09955; 1NSA.
DR MEROPS: M14.010; -.
DR MGD; MGI:88479; Cp33.
DR InterPro: IPR001316; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CARBOXYPTASE.1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT 245 268 BY SIMILARITY.
SO SEQUENCE 417 AA; 48790 MW; A2B30A068D1BA6D CRC64;

```

Query Match 35.2%; Score 632; DB 1; Length 417;
 Best local similarity 38.2%; Pred. No. 1, le-46;
 Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7;

```

OY 1 PGGQVLAALPRTSRQVQLQNTTTEYIVLQPVATADLVKKKQVHFVNASDVNVKA 60
DB 21 FDEKVFYKVLQNEKHASYLKNLTOSIEIDFVYPAIDHIAVMYVDERVSEKESQGIS 80
OY 61 HLNVSGIPCSVLLADVEDLQOQIS-NDVYSPRASASYEQHYSLWEIYSWIFTERHP 119
DB 81 TLEQKHITHEILLHDQLEIKQFDVKEIAGRHNS--YAKYNDMDKIVSYEMKLEKHP 137

```

```

OY 120 DMLTRKHIGSSPEKRYPLVLKVGSGKEQTAKNAIMDGIHAREMISPAFCIMPT----- 173
DB 138 EMVSKRIKIGTIVEDNPLYLKI-GKKDGERKAITMDGSHAREMISPAFCQMFYQATKS 196
OY 174 -GHN-----RMNRKNRSFPANNHCIGTDLNRNFA 202
DB 197 YGKNIMTKLDRMNFYVLPVFNVNGYIWSWTQDRMRKNRSRNQNSTCIGTDLNRNF-D 255
OY 203 KHWEEGASSSCSEETCGLYPESEPEFKAVASFLRRINQIKAYISMHSYQHIVPYS 262
DB 256 VSMSSPNTNPPCLNVRKGPAPSESEKRAVTNPIRHLMSIKAYITPHSYQMLLIPYC 315
OY 263 YTRSKSKDHEELSLVASEAVRAIEKTSNRYTHGSEFTLYLAPGGDDMYIDLGITKS 322
DB 316 YTFKLPRPHQDLKAVARIATDAL-STREYERYITGPIASTIYIKTSGSSLMDWVDLGIKHT 374
OY 323 F 323
DB 375 F 375

```

```

RESULT 5
CBPC_CANFA STANDARD: PRT: 416 AA.
ID CBPC_CANFA
AC P55261;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE membrane associated protein) (ZAP47).
GN CPB1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas.
RA Fukuoaka S.-I.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-Lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D78348; BAA11366.1; -
DR HSSP: P09955; 1PBA.
DR MEROPS: M14.003; -.
DR InterPro: IPR001316; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CARBOXYPTASE.1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 110 416 CARBOXYPEPTIDASE B.
FT METAL 175 175 ZINC (BY SIMILARITY).
FT METAL 178 178 ZINC (BY SIMILARITY).
FT METAL 303 303 ZINC (BY SIMILARITY).
FT ACT_SITE 377 377 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 244 267 BY SIMILARITY.

```

FT DISUFLD 258 272 BY SIMILARITY
SQ SEQUENCE 416 AA: 47595 MW: DFEI992CD052F8AB4 CRC64;

Query Match 35.1%; Score 630.5; DB 1; Length 416;
Best Local Similarity 36.6%; Pred. No. 1,5e-46;
Matches 132; Conservative 64; Mismatches 122; Indels 43; Gaps 6;

OY 1 FOSGOVLAALPRTSRQVOYLQNLTTTTEYLWQPTADLIYKKQVHFFVNASDVNDVKA 60
DB 21 FEGEKEFRVNEDEHNINLHTLASTQTIDFWKPDVSTQIKPHSTADFRKADILTED 80
OY 61 HUNVSGIPCSVLADVEDLIQOQISNDVSPRASASYOYHSLNITYSWIEFIERHPD 120
DB 81 FIKQNELHVEYLINLRLLEQOFGRQV--PATGHSYERKYNMETEIAMTQOVTSENDP 137
OY 121 MLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDICGHAREMISPAFLMFI----- 173
DB 138 LISRSISGTFEGRTIYLKLV-CKAGQNKPAIFMDCGFHAREMISPAFQWVFREXIRTY 196
OY 174 -----GH-----NRMKRKRSEFYANNHCIGTDLNRFASK 203
DB 197 GOEIHTELLDKLDFVLPVGNIDGYVYTWTKRMKRKTRSTQVGTNCVGTDEPTNRF-DA 255
OY 204 HMCCEGASSSGSEFTYCGIYPSSEPEVKAVASFLRNINQIKAYISMHSYQIIVPRYS 263
DB 256 GWCKIGASNNPCDEYCGPAAESEKTKALANFIRSNLSIKAYLIHSHSOMMLYPRYS 315
OY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSEFTLYLAPGGDDMIYDLGIKYSF 323
DB 316 DYKLENNMELNALAKATYKEL-ATLHGTRKYTGPGATTIYPAAGSDMDADOGIKYSF 374
OY 324 T 324
DB 375 T 375

RESULT 6
CBPC_HUMAN STANDARD; PRT; 417 AA.
AC P15088;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
GN CPB3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=lung;
RX MEDLINE=90083291; Pubmed=2594780;
RA Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA Austen K.F., Serafin W.E.;
RT "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT comparison of the protein with mouse mast cell carboxypeptidase A and
RT rat pancreatic carboxypeptidases";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mast cells;
RX MEDLINE=92105393; Pubmed=1729276;
RA Reynolds D.S., Gurley D.S., Austen K.F.;
RT "Cloning and characterization of the novel gene for mast cell
RT carboxypeptidase A";
RL J. Clin. Invest. 89:273-282(1992).
RN [3]
RP SEQUENCE OF 110-417 FROM N.A.
RX MEDLINE=9233165; Pubmed=1629626;
RA Natsuaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuaki M.,

RA Wintrob B.U., Rutter W.J., Goldstein S.M.;
RT "Human skin mast cell carboxypeptidase: functional characterization,
RT cDNA cloning, and genealogy."
RL J. Invest. Dermatol. 99:138-145(1992).
RN [4]
RP SEQUENCE OF 110-137.
RX MEDLINE=89214692; Pubmed=2708524;
RA Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintrob B.U.;
RT "Human mast cell carboxypeptidase. Purification and
RT characterization."
RL J. Clin. Invest. 83:1630-1636(1989).
CC -! CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
CC amino acid.
CC -! SUBCELLULAR LOCATION: Secretory granules.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M27717; AAA5652.1; -
DR EMBL: M73720; AAA59568.1; -
DR EMBL: M73716; AAA59568.1; JOINED.
DR EMBL: M73717; AAA59568.1; JOINED.
DR EMBL: M73718; AAA59568.1; JOINED.
DR EMBL: M73719; AAA59568.1; JOINED.
DR EMBL: S40234; AAB2578.2; ALT_SEQ.
DR PIR: A43929; A43929.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR MIM: 114851; -.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CRBOXYPRASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 15
FT PROPEP 16 109
FT CHAIN 110 417
FT METAL 176 176
FT METAL 179 179
FT METAL 304 304
FT ACT_SITE 378 378
FT DISULFID 173 186
FT DISULFID 245 268
FT SEQUENCE 417 AA; 48700 MW; 8CB90DB758117B24 CRC64;
Query Match 34.6%; Score 622; DB 1; Length 417;
Best Local Similarity 35.5%; Pred. No. 7,8e-46;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;

OY 1 FOSGOVLAALPRTSRQVOYLQNLTTTTEYLWQPTADLIYKKQVHFFVNASDVNDVKA 60
DB 21 FEGEKEFRVNEDEHNINLHTLASTQTIDFWKPDVSTQIKPHSTADFRKADILTED 80
OY 61 HUNVSGIPCSVLADVEDLIQOQISNDVSPRASASYOYHSLNITYSWIEFIERHPD 119
DB 81 ALDQNMHTEILLHDQEEIEKQFDVKEIDPGRHS--YAKYNNWETIVAMTERMDKYP 137
OY 120 DMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDICGHAREMISPAFLMFI----- 173
DB 138 EWSRIKIGSTVEDNPLVYAKI-GEKNERKKAIFMDCGFHAREMISPAFQWVFYQATKT 196
OY 174 -----GH-----NRMKRKRSEFYANNHCIGTDLNRFASK 202

```

Db 197 YGNKTKMTLLDRMNFYIIPEVNDGYIWSWTKNRMKRNKSNKSKICGTDLNRNF-N 255
OY 203 KHWCEGASSSSETCYCGIYPESEPEKAVASFLRNINQIKAVYISMSYQHIYPPS 262
Db 256 ASMNSLPTNNDPCADNYKRSAPSESEKETKAVTNFINSHLNEIKVYITFSYOMLLEPPG 315
OY 263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHSFTLYLAPGGDWIYDGIKYS 322
Db 316 YTSKLPEPNHEDLAKVAKIGTDVIL-STRYETRYIYGPTESTIYPISSSLDMAYDGIKHT 374
OY 323 F 323
Db 375 F 375

```

```

RESULT 7
CBPC_BOVIN STANDARD: PRT: 306 AA.
AC P00732:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B (EC 3.4.17.2).
GN Cpb.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN [2]
RP SEQUENCE OF 31-93: 131-181: 263-265 AND 292-306.
RX MEDLINE=74260705; PubMed=483744;
RA Schmidt J.J., Hirs C.H.W.;
RT "Primary structure of bovine carboxypeptidase B. Inferences from the
RT locations of the half-cysteines and identification of the active site
RT arginine.";
RL J. Biol. Chem. 249:3756-3764(1974).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RP MEDLINE=76265065; PubMed=957425;
RA Schmid M.F., Herliot J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RL J. Mol. Biol. 103:175-190(1976).
RN [4]
RP ACTIVE SITE.
RX MEDLINE=70007159; PubMed=5344132;
RA Plummer T.H. Jr.;
RT "Isolation and sequence of peptides at the active center of bovine
RT carboxypeptidase B.";
RL J. Biol. Chem. 244:5246-5253(1969).
RN [5]
RP ACTIVE SITE.
RX MEDLINE=73061487; PubMed=4565668;
RA Kimmel M.T., Plummer T.H. Jr.;
RT "Identification of a glutamic acid at the active center of bovine
RT carboxypeptidase B.";
RL J. Biol. Chem. 247:7864-7869(1972).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC PIR: A00912; CPOB.
DR PDB: 1CPB; 30-SEP-83.
DR MEROPS: M14.003; -.
DR InterPro: IPR000834; Zn_carpopept.
DR Pfam: PF00246; Zn_carpopept; 1.

```

```

DR PRINTS: PR00765; CROXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT DISULFID 149 163
FT METAL 66 66
FT METAL 69 69
FT METAL 194 194
FT ACT_SITE 246 246
FT ACT_SITE 268 268
SQ SEQUENCE 306 AA; 34612 MW; C329D2655C44A172 CRC64;
NUCLEOPHILE.

```

```

Query Match 32.6%; Score 586; DB 1; Length 306;
Best Local Similarity 43.9%; Pred. No. 6,2e-43;
Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;

```

```

OY 98 YEQYSLNEIYSWIEFIERHHPDMITKHISSEFKYPLVLYKSGKEQTAKNAIWDG 157
Db 6 YEKYNNMETIEMWEOVASENDLISRSAICTTFLGNTIYLLKV-GKRGSKPAVPMDCG 64
OY 158 IHAREWISPAFCLWFI-----GH-----NRMWR 180
Db 65 FHAREWISPAFCOMFVREAVRTYGREIHMEFLDKDFYVLPVYNIDGYITWTNNMR 124
OY 181 KNRSFYANNHICIDLNRFNSKHWCEGASSSSETCYCGIYPESEPEKAVASFLRN 240
Db 125 KTRSTRAGSSCTGTDLNNF-DAGWCSTIGASNNPCSETCYSGSAESESERKAVADFI 183
OY 241 INQIKAVYISMSYQHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHS 300
Db 184 LSLIKAVYITFSYOMLLEPPSYDYKLPKNNVELNTLAKAVKVL-ASLHGTTYSIGPGA 242
OY 301 ETLLAPGGGDWYIDGIKYSFT 324
Db 243 TTYIPASGGSDMAYDGIKYSFT 266

```

```

RESULT 8
CBPC_RAT STANDARD: PRT: 309 AA.
AC P21961:
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE A3).
GN CPB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RX MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., Le Trong H., Woodbury R.C., Walsh K.A.,
RA Neurath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules.";
RL Biochemistry 30:648-655(1991).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC PIR: A33118; A33118.
DR PIR: A38395; A38395.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR InterPro: IPR000834; Zn_carpopept.
DR Pfam: PF00246; Zn_carpopept; 1.

```

Rt	Fanning Z., Koabe B., Stewart C.-B.	Rutler W.J., Goldsmith E.J.;
Rt	"Structural evolution of an enzyme specificity. The structure of rat	
Rt	carboxypeptidase A2 at 1.9 Å resolution."	
RL	J. Biol. Chem. 266:24606-24612(1991).	
CC	-1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A	
CC	(EC 3.4.17.1), but with a preference for bulkier C-terminal	
CC	residues.	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE	
CC	ZINC CARBOXYPEPTIDASE FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isdb.ch/announce/	
CC	or send an email to license@isdb.sib.ch).	
CC	-----	
DR	EMBL; M23721; AAAA40956.1; -	
DR	EMBL; M23714; AAAA40956.1; JOINED.	
DR	EMBL; M23715; AAAA40956.1; JOINED.	
DR	EMBL; M23716; AAAA40956.1; JOINED.	
DR	EMBL; M23717; AAAA40956.1; JOINED.	
DR	EMBL; M23718; AAAA40956.1; JOINED.	
DR	EMBL; M23719; AAAA40956.1; JOINED.	
DR	EMBL; M23720; AAAA40956.1; JOINED.	
DR	EMBL; S79837; -; NOT_ANNOTATED_CDS.	
DR	PIR; A32128; A32128.	
DR	HSSP; P48052; IATE.	
DR	MEROPEPS; M14.002; -	
DR	InterPro; IPR003146; Proper_M14.	
DR	InterPro; IPR000834; Zn_carboxpept.	
DR	Pfam; PF02244; Proper_M14; 1.	
DR	Pfam; PF00246; Zn_carboxpept; 1.	
DR	PRINTS; PR00765; CARBOXYPEPTSEA.	
DR	PROSITE; PS00132; CARBOXYPEPT_ZN.1; 1.	
DR	PROSITE; PS00133; CARBOXYPEPT_ZN.2; 1.	
KW	Hydrolase; Carboxypeptidase; metalloprotease; zinc; zymogen; signal.	
FT	SIGNAL	1
FT	PROPEP	17
FT	CHAIN	112
FT	METAL	173
FT	METAL	177
FT	METAL	180
FT	METAL	180
FT	ACT_SITE	304
FT	DISULFID	378
FT	DISULFID	246
FT	SEQUENCE	318
SO	SEQUENCE	417 AA; 46912 MW; BBDC41A830F2D45 CRC64;
Query Match	30.6%; Score 549.5; DB 1; Length 417;	
Best Local Similarity	34.6%; Pred. No. 1.3e-39;	
Matches 134; Conservative 63; Mismatches 127; Indels 63; Gaps 12;		
OY	1 FQSGVLAALPRTSQOVQLONTLT--YEYLWGPVYADLIYKKQYHFPNASDVNV 58	
Db	20 FVGDDVLITPSHERQIFTLQLLEAEHELDLDFMSPST---IPRETVHVRPFASIAQ 75	
OY	59 KAHLNVGSIPCSVALIADVEDLLQQGISN---DIYSPPRASASYEGQSHSLNETIYSIEFIT 115	
Db	76 KYFLSSGGIIDSIMIEDVOYLLDGRERMLFNQQRKRGNGNFEEAKYHTLLEELIOQMOLIV 135	
OY	116 ERHPDMLTRIHIGSFSEKYPLLYLVKS--GKEOTAKNAIWIIDCGIHAREMWISPAFLM-- 171	
Db	136 AENPGVAISKVNIGSSFENRPKNPVNLKFSGGD---KPATIMDLAGIHAREMVYQTATALMTAN 192	
OY	172 -----FI-----GH-----NRMKRRRSFYANNHGCTDIN 197	
Db	193 KIASDYGTDPALTSLNTLNDLFILFPLVTNPDGIVGESQTNNRMKRTRKRSGSGCYGVDPN 252	
OY	198 RNFAKHWCEBGCASSSCSETEYYCGLYPSEPEPKAVASFLLRNINQIKAYISMHSYOH 257	
Db	253 RNM-DANGGCGACASPSCSDSYHGKRPNSVEVKGIYDFIKSH-GKVAAFTILHSYOLL 310	

QY 258 VPPYSTRSKSDHEELSLVASEAVRAIKRTSKNTRYTHGSEETLYLAPGGDDWYDL 317
 DB 311 MEPPGYKCYKPPDFNELDEVAQKAQAL-KRLHGTSYKVGPCSVIYQASGSIDMAYDL 369
 QY 318 GKXSTSTN-----PVEKLLP 334
 DB 370 GKXSFAPFLRDTAFYGLLPKQILP 396

RESULT 10
 CBPA_BOVIN STANDARD: PRT: 419 AA.

AC 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase A precursor (EC 3.4.17.1).
 GN CPA.
 OS Bos taurus (bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID:9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91151335; PubMed=1998496;
 RA Le Huecrou I., Guilloteau P., Toullec R., Puigserver A., Wicker C.;
 RT "Cloning and nucleotide sequence of a bovine pancreatic
 RT preprocarboxypeptidase A cDNA."
 RL Biochem. Biophys. Res. Commun. 175:110-116(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=96096549; PubMed=8522204;
 RA Goo J.H., Kim K.H., Choi K.Y.;
 RT "Cloning, sequencing and expression of the gene encoding a major
 RT allelic preprocarboxypeptidase A from bovine pancreas."
 RL Gene 165:333-334(1995).
 RN [3]
 RP SEQUENCE OF 111-417.
 RX MEDLINE=71151183; PubMed=5102489;
 RA Bradshaw R.A., Walsh K.A., Neuraath H.;
 RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
 RT chymotryptic peptides of the cyanogen bromide fragment F-1."
 RL Biochemistry 10:938-950(1971).
 RN [4]
 RP REVISIONS TO 138 AND 141.
 RX MEDLINE=72138789; PubMed=5143102;
 RA Petra P.H., Hermanson M.A., Walsh K.A., Neuraath H.;
 RT "Characterization of bovine carboxypeptidase A (Allan)."
 RL Biochemistry 10:4023-4025(1971).
 RN [5]
 RP SEQUENCE OF 17-120.
 RX MEDLINE=89150306; PubMed=3147705;
 RA Wade R.D., Hass G.M., Kumar S., Walsh K.A., Neuraath H.;
 RT "The amino acid sequence of the activation peptide of bovine pro-
 RT carboxypeptidase A."
 RL Biochimie 70:1137-1142(1988).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
 RX MEDLINE=83294519; PubMed=6887246;
 RA Rees D.C., Lewis M., Lipscomb W.N.;
 RT "Refined crystal structure of carboxypeptidase A at 1.54-A
 RT resolution."
 RL J. Mol. Biol. 168:367-387(1983).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
 RX MEDLINE=96003618; PubMed=7556081;
 RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
 RT "The three-dimensional structure of the native ternary complex of
 RT bovine pancreatic procarboxypeptidase A with proproteinase E and
 RT chymotrypsinogen C."

RL EMBL J. 14:4387-4394(1995).
 RN [8]
 RP VARIANT ALLELIC.
 RX MEDLINE=69283620; PubMed=5817619;
 RA Petra P.H., Bradshaw R.A., Walsh K.A., Neuraath H.;
 RT "Identification of the amino acid replacements characterizing the
 RT allelic forms of bovine carboxypeptidase A."
 RL Biochemistry 8:2762-2768(1969).
 CC -1 CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
 CC amino acid.
 CC -1 SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
 CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
 CC PROTEINASE E.
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -1 DATABASE: NAME=worthington-enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M61851; AAA30426.1; -;
 DR EMBL: M61851; AAA30427.1; -;
 DR EMBL: 233906; CAA83955.1; -;
 DR PIR: JN0126; CPBOA.
 DR PIR: A31406; A31406.
 DR PDB: 3CPA; 15-JAN-87.
 DR PDB: 4CPA; 22-OCT-84.
 DR PDB: 5CPA; 15-JAN-87.
 DR PDB: 6CPA; 15-OCT-91.
 DR PDB: 7CPA; 31-JAN-94.
 DR PDB: 8CPA; 31-JAN-94.
 DR PDB: 1CBX; 31-JAN-94.
 DR PDB: 1CPB; 15-OCT-94.
 DR PDB: 2CTB; 31-JAN-94.
 DR PDB: 2CTC; 31-JAN-94.
 DR PDB: 1ARL; 01-AUG-96.
 DR PDB: 1ARL; 17-AUG-96.
 DR PDB: 1BAV; 01-APR-97.
 DR PDB: 1YME; 12-FEB-97.
 DR PDB: 1CPX; 05-AUG-98.
 DR PDB: 1PYT; 27-JAN-97.
 DR MEROPS: M14.001; -;
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxypept; 1.
 DR PRINTS: PR00765; CRBOXYPEPASE.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal;
 KW 3D-structure; Polymorphism.
 FT SIGNAL 1 16
 FT PROPEP 17 110
 FT CHAIN 111 419
 FT METAL 179
 FT METAL 182
 FT METAL 162
 FT METAL 306
 FT ACT_SITE 358 358
 FT ACT_SITE 380 380
 FT DISULFID 248 271
 FT VARIANT 289 289
 FT VARIANT 338 338
 FT VARIANT 415 415
 FT VARIANT 415 415
 FT CONFLICT 95 95
 FT CONFLICT 199 199
 FT CONFLICT 203 203
 FT CONFLICT 224 224
 FT CONFLICT 224 224
 I -> V (IN ALLELIC VARIANT).
 E -> A (IN ALLELIC VARIANT).
 L -> V (IN ALLELIC VARIANT).
 S -> L (IN REF. 5).
 D -> N (IN REF. 3).
 D -> N (IN REF. 3).
 D -> N (IN REF. 3).


```

FT CONFLICT 232 232 Q -> E (IN REF. 3).
FT CONFLICT 295 295 D -> N (IN REF. 3).
FT TURN 114 116
FT TURN 119 120
FT HELIX 125 138
FT TURN 140 142
FT STRAND 143 150
FT TURN 152 153
FT STRAND 156 162
FT STRAND 171 176
FT TURN 180 181
FT HELIX 183 199
FT TURN 200 202
FT HELIX 204 212
FT STRAND 214 218
FT HELIX 223 231
FT TURN 232 232
FT TURN 234 235
FT STRAND 242 242
FT TURN 244 245
FT STRAND 249 249
FT TURN 253 254
FT TURN 261 262
FT STRAND 267 267
FT TURN 270 271
FT TURN 273 274
FT STRAND 275 275
FT TURN 280 281
FT HELIX 284 296
FT STRAND 299 307
FT STRAND 311 314
FT TURN 324 325
FT HELIX 326 344
FT STRAND 349 352
FT HELIX 353 356
FT TURN 357 357
FT HELIX 364 370
FT TURN 371 372
FT STRAND 375 381
FT TURN 388 389
FT HELIX 393 416
SQ SEQUENCE 419 AA; 47082 MM; 21B86407B3BFC452 CRC64;

Query Match 29.8%; Score 536.5; DB 1; Length 419;
Best Local Similarity 34.8%; Pred. No. 1.6e-38;
Matches 129; Conservative 59; Mismatches 124; Indels 59; Gaps 10;

```

```

DB 368 SYNOGIRISFT 378
RESULT 11
ID CBP1_RAT STANDARD; PRT; 419 AA.
AC P00731;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN CPAL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82105986; PubMed=6275388;
RA Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standing D.N.,
RA Picot R.L., Valenzuela P., Rutter W.J.;
RT "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
RT characterization of the gene."
RT Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family."
RL J. Biol. Chem. 263:17837-17845(1988).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: V01232; CA24542.1; -
DR EMBL: J00713; AAA40893.1; -
DR EMBL: M23990; AAA40955.1; -
DR EMBL: M23960; AAA40955.1; JOINED.
DR EMBL: M23985; AAA40955.1; JOINED.
DR EMBL: M23986; AAA40955.1; JOINED.
DR EMBL: M23987; AAA40955.1; JOINED.
DR EMBL: M23988; AAA40955.1; JOINED.
DR EMBL: M23989; AAA40955.1; JOINED.
DR PIR: A00911; CPERA.
DR PIR: B32129; B32129.
DR HSSP: P00730; 1PPT.
DR MEROPS: M14.001;
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept. 1.
DR PRINTS: PR00765; CARBOXYPEPTASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolyase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 182 182 ZINC (BY SIMILARITY).

```

```

RN FEBS Lett. 420:7-10(1997).
RL [1]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RA MEDLINE=98046021; PUBMED=9384570;
RA Garcia-saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT "The three-dimensional structure of human procarboxypeptidase A2.
RT Deciphering the basis of the inhibition, activation and intrinsic
RT activity of the zymogen."
RL EMBO J 16:6906-6913(1997).
CC -1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC residues.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC DR EMBL: D19977; AAA74425.1; -.
CC DR PDB: 1AYE; 13-JAN-99.
CC DR MEROPS: M14.002; -.
CC DR MIM: 600688; -.
CC DR InterPro: IPR001346; Propep.M14.
CC DR InterPro: IPR000834; Zn_CarDopept.
CC DR Pfam: PF002244; Propep_M14; 1.
CC DR Pfam: PF00246; Zn_carDopept; 1.
CC DR PRINTS: PR00765; CARBOXYPTASEA.
CC DR PROSITE: PS00132; CARBOXYPEPT_ZN.1; 1.
CC DR PROSITE: PS00133; CARBOXYPEPT_ZN.2; 1.
CC DR Hydrolase: Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
CC KW 3D-structure.
CC FT SIGNAL 1 16 POTENTIAL.
CC FT PROPEP 17 112 ACTIVATION PEPTIDE.
CC FT CHAIN 113 417 CARBOXYPEPTIDASE A2.
CC FT METAL 177 177 ZINC.
CC FT METAL 180 180 ZINC.
CC FT METAL 304 304 ZINC.
CC FT ACT_SITE 378 378 NUCLEOPHILE.
CC FT DISULFD 246 269
CC FT DISULFD 318 352
CC FT SEQUENCE 417 AA; 46840 MW; 6D8107330A2DBC58 CRC64;

Query Match 28.4%; Score 511.5; DB 1; Length 417;
Best Local Similarity 33.1%; Pred. No. 2.3e-36;
Matches 128; Conservative 61; Mismatches 135; Indels 63; Gaps 10;

QY 1 FQSGGVLAAPPTSPKQYOVQLQNLTTTYEIVL-WQPVTAADLIYVKKQVHFPMASDVQNV 58
DB 20 FVGDQVLEIVPSNENQINLNQLQLEQEHQLDLPFKSP-----TGEETAHAYVPPRYNOAV 75
QY 59 KAHNLVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS--YEQYSLNIEIYSWIEFIT 115
DB 76 KYFLESQGIAYSIMIEDVQLLDKENEMELFNRRRERSGNFAGYHTLEIISQEMDLV 135
QY 116 ERHPMLKRIHIGSFEKXPYLVLKVS-GKQTKAKNAIWIDCGIHARWISPAFCWPIG 174
DB 136 AEHPGLVSKVNIIGSFEKPMVNLKFSTGSD--KPAIMWDAGIHARWYQAALMTATN 192
QY 175 -----HNRMKRNRSPFYANNHIGTDLN 197
DB 193 KIVSDYKDPSTITSLADLIFLPVYTNPDGYVSQTKRNRMRKTRKSVSGSLCVGDPN 252
QY 198 RNFASKHMCCEGASSSCSEFYCGLYPESEPVKAVASFLERRNINQIKAYTISMHSQHI 257
DB 253 RNV-DAGGGGAGGASNPSCSDSYHSPANSSEVEKYSIVDFIKSH-GKAYAFITLHSYQL 310
QY 258 VEPYSYTSKSKHDEELISVSEAVVRAIEKTSKMNRTYTHGSGFTYIAPGCGNDMLYDI 317

```

Db	311	MEPYGYKOTRLDDEDELSEVAQKAQSL-RSLHGKTKKVPICSVIYQASGSIIDMWSDY	369
Qy	318	GIKSYFTSN-----PVEKILP 334	
Db	370	GIKSYFAEFLDRTGRYGLPLPARIOLP 396	
RESULT	13		
CBP1_HUMAN	STANDARD;	PRT;	419 AA.
AC	P15085;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-JUN-1993 (Rel. 26, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Carboxypeptidase A1 precursor (EC 3.4.17.1).		
GN	CPA1 OR CPA.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
NCBI_TaxID	9606;		
FN	(1)		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Pancreas;		
RX	MEDLINE=93038569; PubMed=1417781;		
RA	Catasus L., Villegas V., Pascual R., Aviles F.X.,		
RT	Wicker-Planquart C., Pulgserver A.;		
RT	"cDNA cloning and sequence analysis of human pancreatic		
RL	procarboxypeptidase A1.";		
RL	Biochem. J. 287:299-303(1992).		
RP	(2)		
RP	SEQUENCE OF 17-42.		
RA	MEDLINE=89153096; PubMed=2920728;		
RX	Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;		
RT	"Purification and properties of five different forms of human		
RT	procarboxypeptidases.";		
RL	Eur. J. Biochem. 179:609-616(1989).		
CC	-1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-		
CC	amino acid.		
CC	-1- SUBUNIT: MONOMER.		
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE		
CC	ZINC CARBOXYPEPTIDASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X67318; CAA47732.1; -.		
DR	PIR; S02810; S02810.		
DR	PIR; S29127; S29127.		
DR	HSSP; P00730; 1PPT.		
DR	MEROPS; M14.001; -.		
DR	MIM; 114850; -.		
DR	InterPro: IPR003146; Propep_M14.		
DR	InterPro: IPR000834; Zn_carboxypep.		
DR	Pfam; PF02244; Propep_M14; 1.		
DR	Pfam; PF00246; Zn_carboxypep; 1.		
DR	PRINTS; PR00765; CRBOXYPTASEA.		
DR	PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.		
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.		
KW	Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.		
FT	SIGNAL	1	16
FT	PROPEP	17	110
FT	CHAIN	111	419
FT	METAL	179	179
FT	METAL	182	182
FT	METAL	306	306
FT	ACT_SITE	358	358
FT	ACT_SITE	380	380
FT	NUCLEOPHILE (BY SIMILARITY).		
FT	ACT_SITE	380	380
FT	NUCLEOPHILE (BY SIMILARITY).		

```

FM DISULFID 248 271 BY SIMILARITY.
SQ SEQUENCE 419 AA; 47140 MW; 439FAFFFAE958B1 CRC64;

Query Match 27.4%; Score 493.5; DB 1; Length 419;
Best Local Similarity 33.1%; Pred. No. 7.9e-35;
Matches 126; Conservative 58; Mismatches 118; Indels 79; Gaps 12;

QY 1 FQSGOVLALPRTSRQOVYLONLTTYEIVL--WQ-----PYTADLIYKKQVHEFVN 51
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20 FVGHQVLRISVADEAKOVQKKELEDLEHLDLFMGPAHPGSPIDV-----RVPF--- 69
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 52 ASDVDNVAHLNVSGIPCSVLLADVEDL---QQOI-----SNDVSPRASASYEQ 100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 -PSIOAVAFIEFHSGHSIYETIMIEDVQSLDEEQMEAFRRANSTDFN-----YAT 121
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 101 YHSLNEIKSWIEFTEFRRPDMLTRKHSSSEKPYLYLKYSGKEQTAKNAIWIDCGIHA 160
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 YHTLEIYDFPDLVAENPHLVSKIQIGNTYEGRIYLVKFS-YQGSRRPAIWDITGHS 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 161 REMISPAFCWFI-----GHNRRMRNR 183
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 REMVQAGSVMPFAKKITQDYGODAAFTAILDITLDFIEIYVNPDPGAFHTSTNRMRKTR 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 184 SFYANNHCITGDLNRNFSKMCBEGASSSCSEFYCYLLYBSEPEVAAVAFPLRNINO 243
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 SHTAGSLCIGVDPNNMVGAFGL-SGASNNCSCEIYFHKFNSEVEYISYDFVKGHCN- 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 244 IKAYSIMSGISOHYEFPSPYTRSKSDHEELSLVSEVRAIEKTSKTRTYHGHGSETL 303
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 299 IKAFLSIHSYGLMLPPIGTYKTEPPDPDELDQLSKAAVTLAL-ASLYTKFNYSIIKAI 357
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 304 YLAPGGDDWIYDLGIKYSFT 324
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 358 YQASGSTIDWYYSQGIKYSFT 378
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
CBP4_HUMAN STANDARD; PRT; 421 AA.
ID CBP4_HUMAN
AC Q9UII42:
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN CPA4 OR CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP
RX MEDLINE=99310537; PubMed=10383164;
RA Huang H., Reed C.P., Zhang J.S., Shidhar V., Wang L., Smith D.I.;
RT "Carboxypeptidase A3 (CPA3):: a novel gene highly induced by histone
RT deacetylase inhibitors during differentiation of prostate epithelial
RT cancer cells.";
RL Cancer Res. 59:2981-2988(1999).
CC -! FUNCTION: Could be involved in the histone hyperacetylation
CC pathway.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF095719; AAF23230.1; -.
CC HSSP; P48052; IAYE.

```

```

DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF00244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carboxypept; 1.
DR PRINTS: PR00765; CARBOXYPEPASE.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 113
FT CHAIN 114 421
FT METAL 181 181
FT METAL 184 184
FT METAL 308 308
FT ACT_SITE 382 382
FT DISULFID 250 273
SO SEQUENCE 421 AA; 47379 MW; 91883D25B705B6CA CRC64;

Query Match 27.38; Score 491; DB 1; Length 421;
Best Local Similarity 32.7%; Pred. No. 1.3e-34;
Matches 123; Conservative 59; Mismatches 126; Indels 68; Gaps 11;

QY 1 FGGGVALALPRTSRQVQLNLTTEIYL--WQVTA-----DLVKKKQVHEFNAS 53
DB 21 FFGDQVLRIVNRGDELSKLSQLYNSNNLKNTEKSSFRNPVAVLPVSLDAF---- 76
QY 54 DVNKAHLNVSGIPCSVLADVEDLI--QQQISNDVSPRASAY-YEQYHSLNETYS 109
DB 77 -----KSLNSQGLEAVVTEDLQALLDNDEDEQHNQGEGRSSNNNYGAYHSLDAIYH 131
QY 110 WIEFTRHDMLTKIIGSFERYPLVYLKVSKEQTAANKAIWIDGIAREKISPAFC 169
DB 132 EMDNIADFPDLARVYKIGHSFENRPMVYLFKSTKGVRRPAVNLNGHSREVISQATA 191
QY 170 LW-----FI-----GHNRMKKNSEFFANNHCI 192
DB 192 IWRARKVSDYQDRPATSLIEKMDIFLRYANPDGYVYTOTNRKMRKTRSRPGSSCI 251
QY 193 GTDLNRN---FASKHWCCEGASSSCSEFYCGLYPESEPEVKAASFLLRNINOIKAYI 248
DB 252 GADPNRMNMSAFGK-----GASDNPCESEVYHGPHANSEVEKSYVDIQLHGN-FKGEI 305
QY 249 SMSISQHIPEPYSTRSKSKDHELSLVASEAVRALEKTSKNTRYTHGHSETLAPG 308
DB 306 DLHSYSQLLMYPGYGYKKAADAEELDKVAKLAKALASVS-GTEYOVGPTCTTVYPSG 364
QY 309 GGDDMIYDLGIKYSFT 324
DB 365 SSIDMAYDNIGIKFAFT 380

RESULT 15
CBP2_SIMVI STANDARD; PRT; 304 AA.
AC P42788;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=gut;
RX MEDLINE=9403864; PubMed=8269093;
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins.";
RL Insect Mol. Biol. 1:149-163(1993).

```

```

CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L08481; AAA18531.1; -.
DR HSSP: P48052; IAYE.
DR InterPro: IPR000834; Zn_carboxypept.
DR Pfam: PF00246; Zn_carboxypept; 1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc.
FT NON_TER 1 1
FT METAL 58 58
FT METAL 61 61
FT METAL 184 184
FT ACT_SITE 236 236
FT ACT_SITE 259 259
FT DISULFID 125 148
FT DISULFID 125 148
SO SEQUENCE 304 AA; 34849 MW; 26E3F8A6AA9144 CRC64;

Query Match 22.98; Score 411.5; DB 1; Length 304;
Best Local Similarity 33.38; Pred. No. 5.3e-28;
Matches 94; Conservative 52; Mismatches 85; Indels 51; Gaps 9;

QY 100 QYHSLNEIYSWIEFTRHDMLTKIIGSFERYPLVYLKVSKEQTAANKAIWIDGIIH 159
DB 1 QYHSLNEIYSWIEFTRHDMLTKIIGSFERYPLVYLKVSKEQTAANKAIWIDGIIH 159
QY 160 AREWISPA---FCL-----WFI-----GH-----NRMKRNS 184
DB 59 AREWITATATYLLNELTSKNTIREMAENYDVIPTNPDDGYVYTHTDNRMRKTRS 118
QY 185 FYANHCIGTDLNRNPAKSHWCCEGASSSCSEFYCGLYPESEPEVKAASFLLRNINOI 244
DB 119 PNPSLCACTDPNNRW-NFHMDEGTSRPTETLYGKKAASEVETNSFSFLTLGQI 177
QY 245 KAYISMSQHIPEPYSTRSKSKDHELSLVASEAVRALEKTSKNTRYTHGHSETLAPG 304
DB 178 KVTYLAHSYSQLLLPFGYHTQHTYHNDLQALGDAARSLAQ-RYGDYTVGNITDAIY 236
QY 305 LAPGGDDMIY---DLGIKYSFTSNP-----PYEKLLP 334
DB 237 PASGGSMDMAYDLDIPIAYTELRPRDGMNGFOLPANQITP 278

```

Search completed: September 18, 2002, 16:47:50
Job time: 282 sec

OK of: US-09-980-881-3 to: N_Geneseq_032802:* out_format : pfs
Date: Sep 18, 2002 7:02 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh
-O=/cgat2_1/USPTO.spool/US09980881/rnat.16092002_140130_10714/app_query.fasta.1.1126
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=ptc
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CGN1_1.0 -NCPU=6 -ICPU=3 -LONLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPRY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-3
Query length: 338
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 386.310000

score_list:

```
Sequence      Strd Orig      ZScore      EScore      Len      Documentation
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62802 + 1785.00 3499.84 1.4e-186 15
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV74302 + 1732.00 3397.42 6.9e-181 12
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT1671 + 1732.00 3393.73 1.1e-180 17
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62846 + 1732.00 3393.73 1.1e-180 17
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ41001 + 1727.00 3383.84 3.9e-180 17
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAT8005 + 902.00 1753.89 2.4e-89 14
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ90600 + 646.50 1249.94 2.8e-61 12
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42494 + 637.50 1231.68 2.9e-60 12
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV41795 + 637.50 1231.68 2.9e-60 12
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42506 + 637.50 1231.68 3.0e-60 12
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ03837 + 619.50 1195.63 3.0e-58 12
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV17331 + 619.50 1189.89 6.3e-58 21
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT35760 + 610.00 1180.85 2.0e-57 92
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ90601 + 582.00 1125.52 2.4e-54 92
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ03839 + 571.50 1095.03 1.2e-52 2
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6477 + 569.00 1097.47 8.8e-53 1
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH6479 + 569.00 1097.47 8.8e-53 1
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH6476 + 569.00 1095.43 1.1e-52 1
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH6478 + 569.00 1095.43 1.1e-52 1
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42500 + 567.00 1094.89 1.2e-52 99
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42497 + 567.00 1084.28 1.3e-52 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAAT2804 + 560.00 1080.43 7.8e-52 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42511 + 560.00 1080.36 7.9e-52 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62787 + 560.00 1080.36 7.9e-52 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42512 + 559.00 1078.39 1.0e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62788 + 559.00 1078.39 1.0e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62806 + 556.00 1072.45 2.2e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62790 + 554.00 1068.49 3.6e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62801 + 554.00 1068.49 3.6e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ05477 + 554.00 1067.04 4.4e-51 1
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62789 + 553.00 1066.51 4.7e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62791 + 551.00 1064.53 6.0e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62805 + 551.00 1062.55 7.7e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62803 + 551.00 1062.55 7.7e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62804 + 551.00 1062.55 7.7e-51 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62808 + 549.00 1058.60 1.3e-50 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62800 + 549.00 1058.60 1.3e-50 10
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAV17322 + 549.00 1052.01 3.0e-50 18
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV41819 + 549.00 1052.01 3.0e-50 18
```

/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62802 + 548.00 1056.62 1.7e-50
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62807 + 546.00 1052.66 2.8e-50
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ03838 + 534.50 1030.00 5.0e-49
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAAT1587 + 527.50 1014.04 3.9e-48
/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAQ05476 + 527.50 1013.58 4.1e-48

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC81962

seq_documentation_block:

ID AAC81962 standard; CDNA: 1573 BP.
XX
AC AAC81962;
XX
DT 01-MAR-2001 (first entry)
XX
XX Human brain carboxypeptidase B cDNA.
DE
XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; neurotrophic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX
XX Homo sapiens.
OS
XX Key Location/Qualifiers
XX CDS 18..1100
XX FT /tag= a
XX FT /product= "carboxypeptidase B"
XX PN WO20006717-A1.
XX PD
XX 09-NOV-2000.
XX 01-MAY-2000; 2000MCO-JP02878.
XX 30-APR-1999; 99JP-0125169.
XX (MATS/) MATSUMOTO A.
XX Matsumoto A;
XX WPI, 2000-687534/67.
XX P-PSDB; AAB11457.
XX Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX Claim 2b; Page 64-68; 84pp; Japanese.
XX This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC neurotrophic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
XX Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;
SO
alignment_scores:
Quality: 1785.00 Length: 338
Ratio: 5.297 Gaps: 0
Percent Similarity: 99.704 Percent Identity: 99.112
alignment_block:
US-09-980-881-3 x AAC81962 ..
Align seg 1/1 to: AAC81962 from: 1 to: 1573

```

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVal 17
   |||||
84 TTCAGAGTGGCCAGTTCAGTGGCTTCCATGAAACCTCTAGGCAAGT 133
17 LgInValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTyrGlnP 34
   |||||
134 TCAGAGTTCACAGAAATCTTACTCAACATATGATGATGTTCTCTGCGACG 183
34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
   |||||
184 CGGTACAGCTGACCTTATTGTGTAAGAAAAACAGCTCCATTITTTTGTG 233
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly11 67
   |||||
234 AATGCATTCATGTCGACAAATGCAAGCCCAATTAAATGTGAGCGGAAT 283
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln1 84
   |||||
284 TCCATCGAGTGTCTGTGCGACAGCTGGAGAGATCTTATTCAACACAGCA 333
84 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
   |||||
334 TTTCACAGACACAGTACAGCCCCGAGCCTCCGATCGTACTATGACAG 383
101 TyrHisSerLeuAsnGluIleTyrSerTyrPileGluPheIleThrGlu1 117
   |||||
384 TATCACTCTCAATCAATCAATCTATCTGTGATGAAATTTATTAACGAGAG 433
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluTyr 134
   |||||
434 GCATCTGTATGCTTACAAAATCCACATGTGATCTCTATTGAGAAAT 483
134 yProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
   |||||
484 ACCCACTATGTTTAAAGTTTCTGGAAAGAACAAACACCAAAAT 533
151 AlaIleThrPileAspCysGlyIleHisAlaArgGluTyrPileSerPro1 167
   |||||
534 GCCATATGATGATGACTGGAATCCATGCCAGAGATGATCTCTCTCC 583
167 aPheCysLeuTyrPheIleGlyHisAsnArgMetThrPheGlyAsnArgS 184
   |||||
584 TTTCGCTGTGTTGTCATAGGCCATATCGAATGTGGAGAAACCCGTT 633
184 ePheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
   |||||
634 CTTTCTATGCGAAACATCATTCATCGAGAACACCTGAAATGACAACTTT 683
201 AlaSerLysHisIleTyrCysGluGlnLysLysSerSerSerCysSerG1 217
   |||||
684 GTCCTCAAAACACTGCTGTGAGGAGAGTCAATCCAGTTCCTCATGCTCGA 733
217 uThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValA 234
   |||||
734 AACCTACTGTGACTTATTCCTGAGTCAAGAACCAAGATGAAAGCAAGCG 783
234 LaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMet 250
   |||||
784 TAGAGTTCTTAGAAGAAATATCAACCCAGATTAAAGCATACATCAGCATG 833
251 HisSerTyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLy 267
   |||||
834 CATTCATATCTCCACAGCATATATGTTCTTCATATTCCTATACACGAAAGTAA 883
267 sSerLysAspHisGluGlnLeuSerLeuValAlaSerGluAlaValArgA 284
   |||||
884 AAGCAAGAACCATGAGAACTGTCTCTAGTAGCCAGTCAACAGTCTGCTG 933
284 LaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
   |||||
934 CTATTGACAAATAGTAAATAATACAGGTATACATGAGCCATGGCTCA 983

```

```

301 GluThrLeuTyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLe 317
   |||||
984 GAACCTTATACCTACCTCTCGAGGTGGGACAGATTCGATCATGATTT 1033
317 uGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuP 334
   |||||
1034 GGGCATCAAAATATTCGTTTACATCAACACCACCTGTAGAGAAAGCTTTTGC 1083
334 roLeuSerLeuLys 338
   |||||
1084 CGCTGCTCTGTAATAA 1097

```

seq_name: /SID51/gcdata/geneseq/geneseqn-emb1/NA1999.DAT.AAV74302

seq_documentation_block:

ID AAV74302 standard; cDNA; 1272 BP.

AAV74302;

28-APR-1999 (first entry)

Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.

Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;

polymorphism detection; thrombotic disease; ds.

Homo sapiens.

W09855645-A1.

10-DEC-1998.

02-JUN-1998; 98WO-EP03244.

03-JUN-1997; 97US-0869057.

(SCHD) SCHERING AG.

Morser MJ, Nagashima M;

WPI: 1999-045800/04.

P-PSDB; AAW92270.

Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 humans

Example 1; Page 24; 35pp; English.

This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
 hPCPBthr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphs of the protein
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.

Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

alignment_scores:

Quality: 1732.00

Ratio: 5.124

Percent Similarity: 86.005

Percent Identity: 85.751

alignment_block:

US-09-980-881-3 x AAV74302 ..

Align seg 1/1 to: AAV74302 from: 1 to: 1272

```

1 PheGlnSerGlyGlnValIleuAlaIleuProArthrSerArgGlnVal 17
  |||
67 TTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTTGAACCTCTTAGGCAAGT 116
  |||
17 LglnValIleuGlnAsnLeuThrThrThrThrThrThrThrThrThrThr 34
  |||
117 TCAGATTCTACAGAACTTCTACAACTATGATGATGCTCTCGGAGC 166
  |||
34 roValThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
  |||
167 CGGTAAACAGTGAACCTTATGTGAGAAAAACAGCCATTTTGTGTA 216
  |||
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyI 67
  |||
217 AATGCACTCTGATGTCCGAAATGTAAGCCCATTTAATGTGAGCGGAGT 266
  |||
67 eProCysSerValIleuLeuAlaAspValGluAspLeuIleGlnGlnI 84
  |||
267 TCCATGCAAGTCTGTGCTGGCAGACGTGGAAGATCTTATTCACACAGCA 316
  |||
84 IeSerAsnAspThrValSerProArthrAlaSerAlaSerTyrTyrGln 100
  |||
317 TTTCACACGACACAGTCAGCCCGGAGCCCTCCGATCTACTATGAGCAG 366
  |||
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 117
  |||
367 TATCACTCACTAAATGAAATCTATCTTGATGAAATTTATTAAGTGAAG 416
  |||
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
  |||
417 GCATCCGATGATGCTTACAAATAATCCACATTTGATCTCATTTGAGAGT 466
  |||
134 yProLeuTyrValIleuLysValSerGlyLysGluGlnThrAlaLysAsn 150
  |||
467 ACCCACTCTATGTTTAAAGTTCTGGAAGAAACAAAGCAGCAAAAT 516
  |||
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
  |||
517 GCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 566
  |||
167 aPheCysLeuTrpPheIleGlyHis..... 175
  |||
567 TTCTGCTGTGTTGATCATAGCCATATACTCAATTCATGGATATAG 616
  |||
175 ..... 175
  |||
617 GCCAATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATGCCGCTG 666
  |||
176 ..... AsnArgMetTrpAr 180
  |||
667 GTTAATGTGACGTTATGACTCTCATGGAATAAGATGATGAG 716
  |||
180 gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 197
  |||
717 AAGAACCGTTCTTCTATGCAACATCATGATGATGGAAGACACTGA 766
  |||
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSer 213
  |||
767 ATTAGGAACCTTGTCCAAACACTGTGTGAGGAAGTGTGATCCAGTTTC 816
  |||
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGlu 230
  |||
817 TCATGCTCGAAACCTTACTGTGACTTATCTGAGTCAAAACACAGAAAT 866
  |||
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaT 247
  |||
867 GAAGCAGTGGCTAGTTCTTGAAGAAATATCAACACATTAAGCAT 916
  |||
247 yTrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263

```

```

|||||
917 ACATCAGCATGCAATTCATCTACCCAGCATATAGTTCATATCTCTAT 966
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
  |||
967 ACAGAAAGTAAAGCAAGACATGAGGAACTGCTCTTAGAGCAGATGA 1016
280 uAlaValAlaAlaIleGluLysThrSerLysAsnThrArgTyrThrHis 297
  |||
1017 AGCAGTTCGTGCTTATGAGAAACTAGTAAATAATCCAGTATACATG 1066
297 LysHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrp 313
  |||
1067 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGAGCAT 1116
314 IleTyrAspLeuGlyIleLysTyrSer Phe..... 323
  |||
1117 ATCTATGATTTGGCATCAATATTCGTTTCAATTTGAACTTCAGATAC 1166
324 .....ThrSerAsnProProVal 329
  |||
1167 GGGCACAATACGATTTCTTGTGCCGGAAGCTTACATCAAAACCACTGTA 1216
330 GluLysLeuLeuProLeuSerLeuLys 338
  |||
1217 GAGAGCTTTTGGCCGCTGTCTCTTAAA 1243
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA196.DAT:AA11671
seq_documentation_block:
ID AA11671 standard; DNA; 1749 BP.
XX
AC AA11671;
XX
DT 12-APR-1996 (first entry)
XX
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 41..1312
FT /tag= a
FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT /tag= b
FT mat_peptide 107..1309
FT /tag= c
XX
PN US5474901-A.
XX
PD 12-DEC-1995.
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
XX
PA (GENTECH ) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
DR WPI; 1996-039508/04.
XX
DR P-PSDB; AAR90293.
XX
PT Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hPCPB for use in treating clotting disorders e.g.
PT haemophilia A

```

XX Disclosure: Figure 4: 40pp: English.

CC An antibody which specifically binds human plasma carboxypeptidase B
 CC (hPCPB) and does not cross react with other carboxypeptidases is
 CC useful for the detection of hPCPB in vitro. The antibody is also
 CC used for purifying hPCPB from a sample. Purification comprises
 CC passing a sample thought to contain hPCPB over either a column to
 CC which antibody has been bound, or a plasminogen affinity column,
 CC eluting the column and then recovering the fraction containing the
 CC hPCPB.

XX Sequence 1749 BP: 521 A; 361 C; 342 G; 525 T; 0 other:

Alignment_scores:

Quality: 1732.00 Length: 393
 Ratio: 5.124 Gaps: 3
 Percent Similarity: 86.005 Percent Identity: 85.751

Alignment_block:

US-09-980-881-3 x AAT11671 ..

Align seg 1/1 to: AAT11671 from: 1 to: 1749

```

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVa 17
107 TTTCAAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTTAGCGCACT 156
17 LcInValLeuGlnAsnLeuThrThrThrGlyGluLeuValLeuTrpGlnP 34
157 TCAAGTCTACAGAACTTACTACAAACATATGAGATGTTCTCTGCGACG 206
34 roValThrAlaAspLeuLeuValLysLysLysGlnValHisPhePheVal 50
207 CGGTAAACACCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTGTGTA 256
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyL 67
257 AATGCATCTGATGTCGACAAATGTGAAAGCCCATTTAAATGTCAGCGGAA 306
67 eProCysSerValLeuLeuAlaAspValGluAspLeuLeuGlnGlnLni 84
307 TCCATGCACTGCTGCTGCGGAGAGTGAAGATCTTATTCAACAGCAGCA 356
84 lAsSerAsnAspThrValSerProAlaGlnAlaSerAlaSerTyTrpGluGln 100
357 TTTTCAAGCAGCAGCTGACGCCCCGAGCCTCCGCACTGCTACTATGAACAG 406
101 TyrHisSerLeuAsnGluLeuTySerTrpIleGluPheIleThrGluArg 117
407 TATCACTCACTAAATGAATCTATCTTGATGATGAAATTTATTAAGTGAAG 456
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysT 134
457 GCATCTCATATGCTTACAAAAATCCACATTCGATCCATTTGAGAGT 506
134 yrProLeuTyValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
507 ACCCCTCTATGTTTTAAAGTTTCTGGAAGAAACAAACCCCAAAAT 556
151 AlaIleTrpIleAspCysGlyLysIleHisAlaArgGluTrpIleSerProAl 167
557 GCCATATGATGATGACTGTGAAATCATGCCAGAGAAATGATCTCTCTGCG 606
167 aPheCysLeuTrpPheIleGlyHis..... 175
607 TTTTGGCTTGTGTTCAATGAGCATATCAATTCATTTATGAGGATATAG 656
175 ..... 175
657 GGCATATATACCAATCTCTGAGAGCTTGTGATTTCTATGTTATGCCGCTG 706

```

```

176 .....AsnArgMetTrpArg 180
707 GTTAATGTGAGCGTTATGACTTACTATCATGAAAAAAGAAATCGAAATGTGGAG 756
180 gLysAsnArgSerPheThrAlaAsnAsnHisCysIleGlyThrAspLeuA 197
757 AAGAACCCCTTCTTCTATGCGAACAATCATTCATTCAGGAAACAGACCTGA 806
197 snArgAsnPheAlaSerLysHisIleTrpCysGluGluGlyLysSerSer 213
807 ATTAGAATCTTGCTTCCAAACACTGTGTGAGAGAGAGTCATCCAGTTC 856
214 SerCysSerGluThrTyTrpCysGlyLeuTyTrpGluSerGluProGluVa 230
857 TCATGCTCGAAACCTACTGTGACTTTCCTGAGTGAAGACCAAGAGT 906
230 lLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaT 247
907 GAAGCGAGTGGCTAGTTCCTGAGAAATATCAACACGATTTAAAGCAT 956
247 yrIleSerMetHisSerTySerGlnHisIleValPheProTrpSerTy 263
957 ACATCAGCATGCTATCATCTCCAGCATATAGTGTTCATATATTCCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
1007 ACACGAAAGTAAAGCAAGACCATGAGGAATGCTCTCTAGTACCAAGTGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyThrHisG 297
1057 AGCAGTCTGCTATTGAGAAACTAGTAAATATCAAGGATATACACATG 1106
297 lHisGlySerGluThrLeuTyTrpLeuAlaProGlyGlyLysAspAspTrp 313
1107 GCCATGGCTCAGAAACCTTATACCTAGCTCCGGAAGTGGGAGCATTCG 1156
314 lIleTyAspLeuGlyLysLysTySer.Phe..... 323
1157 ATCTATGATTTGGGCATCAAAATATTCGTTTAACTGAATTCGAGATAC 1206
324 .....ThrSerAsnProProVal 329
1207 GGGCAGATACGATTCCTTGCCGAGCGGAGCTTACATCAAAACCACTGTA 1256
330 GluLysLeuLeuProLeuSerLeuLys 338
1257 GAGAGCTTTGGCCGCTGCTCTTAAAA 1283

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT62846
seq_documentation_block:
ID_AAT62846 standard; DNA: 1749 BP.
XX
AC_AAT62846;
XX
DT_08-MAY-1997 (first entry)
XX
DE_Human plasma carboxypeptidase B coding sequence.
XX
KW_Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
XX
OS_Homo sapiens.
XX
FH_Key
FT_CDS 41..1312 Location/Qualifiers
FT /tag= a
FT /product= Human PCPB
FT sig_peptide 41..106
FT mat_peptide 107..1309
FT /tag= b
FT /tag= c 134..177
FT misc_binding

```



```

FT      /*tag= d
XX      /bound_moelty= 46_bp_probe
PN      US5593674-A.
XX      14-JAN-1997.
PD      XX
XX      01-FEB-1991; 91US-0649591.
XX      01-FEB-1991; 91US-0649591.
XX      01-FEB-1991; 91US-0649591.
PR      14-OCT-1992; 92US-0959944.
PR      15-DEC-1993; 93US-0167727.
PR      19-JUL-1994; 94US-0277540.
PR      27-APR-1995; 95US-0430787.
XX      (GPTH ) GENENTECH INC.
XX      Drayna DF, Eaton DL;
XX      WPI; 1997-099413/09.
XX      P-PSDB; AAMI4733.
XX      Using human plasma carboxypeptidase B in blood coagulation - is
PT      functionally related to carboxypeptidase A and pancreas
PT      carboxypeptidase B
XX      Example 2; Column 37-42; 39pp; English.
XX      This sequence encodes human plasma carboxypeptidase B (PCPB) which
CC      has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC      PCPB may be used therapeutically in haemostatic regulation. PCPB is
CC      purified from human plasma or by transformed cell culture by
CC      extraction using plasminogen bound to a solid phase.
XX      Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other:

alignment_scores:
Quality: 1732.00 Length: 393
Ratio: 5.124 Gaps: 3
Percent Similarity: 86.005 Percent Identity: 85.751

alignment_block:
US-09-980-881-3 x AAT62846 ..
Align seg 1/1 to: AAT62846 from: 1 to: 1749

1 PheGlnSerGlyValLeuAlaAlaLeuProArgThrSerArgGlnA 17
107 TTTCAGAGTGGCCAGTCTAGCTGCTTCTCTAGAACCTTAGGCNAAGT 156
17 LcInValLeuGlnAsnLeuThrThrThrTyrgluIleValLeuTyrglnP 34
157 TCAAGTTTACAGAAATCTTACTACACAAATATGATGTTCTCTGGCAGC 206
34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
207 CGGTAACTGACCTTATTTGCAAGAAAAACAAAGTCCATTTTGTGTA 256
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyI 67
257 AATGCATCTGATGTCACAAATGCAAGCCCATTTAAATGTGACGGAAT 306
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGln 84
307 TCCATGCACTGCTCTGCTGCAAGAGTGAAGATCTTATTCAACACAGAG 356
84 lSerAsnAspThrValSerProArgAlaSerAlaSerTyrgluGln 100
357 TTTTCCAAACACACAGTCAAGCCCCGAGCCTCCGCATCTACTTGAACAG 406
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr 117
|||||

```

```

407 TATCACTCACTAAATGAATCTATTCTTGCATGCAATTTTAACTGACAG 456
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluIuysr 134
457 GCATCCCTGATATGCTTACAAAAATCCACATTTGGATCTCTCATTTTGAGAGT 506
134 yrProLeuTyValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
507 ACCCACTCTATGTTTAAAGTTTCTGAAAAAGAACAAACAGCCCAAAAAT 556
151 AlaIleThrPheLeuPheCysGlyIleHisAlaArgGluTrpIleSerProAl 167
557 GCCATATGATTTGACTGTGGAATCCATGCCAGAAATGATCTCTCTGC 606
167 aPheCysLeuTrpPheIleGlyHis..... 175
607 TTTCTGCTGTGTTGTCATAGCCCATATATCAATCTATGCGATATAG 656
175 ..... 175
657 GGCAATATACCAATCTCCCTGAGGCTTGTGATTTCTATTATGCCGCTG 706
176 ..... AsnArgMetTrpAr 180
707 GTTAATGTGACGCGTTATGACTCTCATGAAAAAAGATGATGTGAG 756
180 gLysAsnArgSerPheTyraAlaAsnAsnHisCysIleGlyThrAspLeuA 197
757 AAGAACCCTGCTTCTTATGCGAACAATCATTTGCTGGAACAGACTGA 806
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 213
807 ATAGGAACCTTGCTCCAAACACTGTGTGAGGAGAGCTGATCCAGTCTCC 856
214 SerCysSerLeuThrTyrglyLeuTyrgluSerProGluSerIuProGluA 230
857 TCATGCTCGAAGACCTTACTGTGGACTTATCTGAGTCAAGACCAAGAGT 906
230 lLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaT 247
907 GAAGGACGTGGCTAGTTTCTTGAGAAAGAAATATCAACACATTAAGCAT 956
247 yrIleSerMetHisSerTyrglySerGlnHisIleValPheProTyrgly 263
957 ACATACACATGATCATCTACTCCAGCATATAGTGTTCATATATCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
1007 ACACGAAGTAAAGACAAAGACATGAGAACTGTCTCTAGTACGCACTGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrglyHisG 297
1057 AGCAGTTCGTGTATGAGAAACTACTTAAATAATACAGATATACACATG 1106
297 LysHisGlySerGluThrLeuTyrlleuAlaProGlyGlyLysAspTrp 313
1107 GCCATGCTCTAGAAACCTTATACCTACTCTCTGAGGTGGAGAGATGG 1156
314 lIleTyraSpLeuGlyIleLysTyrglyPhe..... 323
1157 ATCTATGATTTTGGCATCAATATCTGTTTCAATTAACCTTGCAGATAC 1206
324 ..... ThrSerAsnProProVal 329
1207 GGGCACATACGAGATTTCTGCTGGCGAGCGTTACATCAAAACCCAGTGA 1256
330 GluLysLeuLeuProLeuSerLeuLys 338
1257 GAGAAAGCTTTTGGCCGCTGCTCTTAAAA 1283
seq_name: /sids1/gcgsdata/geneseq/geneseqn-emb1/NA1993.DAT:AA041001
seq_documentation_block:

```

```
ID      AAQ41001 standard; cDNA; 1749 BP.
XX
XX      AAQ41001;
XX
XX      24-AUG-1993 (first entry)
XX
XX      Human plasma carboxypeptidase B gene.
XX
XX      PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
XX      Homo sapiens.
XX
XX      key      Location/Qualifiers
XX      sig_peptide 41..106
XX      mat_peptide 107..1312
XX      misc_feature 134..177
XX      /lag- a
XX      /lag- b
XX      /product- PCPB
XX      /note- "used to obtain full-length clones"
XX
XX      US5206161-A.
XX
XX      27-APR-1993.
XX
XX      01-FEB-1991; 91US-0649591.
XX
XX      01-FEB-1991; 91US-0649591.
XX
XX      (GETH ) GENENTECH INC.
XX
XX      Drayna DT, Eaton DL;
XX
XX      WPI; 1993-151724/18.
XX
XX      P-PSDB; AAR36273.
XX
XX      New human plasma carboxypeptidase B - used as hemostatic
XX      regulator for clotting blood, partic. for treating blood clotting
XX      disorders, e.g. hemophilia
XX
XX      Disclosure; Fig 4; 40pp; English.
XX
XX      Human plasma carboxypeptidase B was isolated from human plasma and
XX      partially sequenced. Oligonucleotide primers were designed based on
XX      the partial amino acid sequences. The primers were used in a PCR
XX      amplification to identify cDNA encoding PCPB from a human liver cDNA
XX      library. The PCR product was capable of encoding the first 37 amino
XX      acids of PCPB; a 46mer probe was used to obtain the full-length
XX      sequence which, although disclosed in the specification, is not
XX      claimed. PCPB inhibits the enzymatic conversion by tPA of
XX      plasminogen to plasmin in the presence of fibrinogen.
XX
XX      Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;
XX
XX      alignment_scores:
XX      Quality: 1727.00      Length: 393
XX      Ratio: 5.125      Gaps: 3
XX      Percent Similarity: 85.751      Percent Identity: 85.496
XX
XX      alignment_block:
XX      US-09-980-881-3 x AAQ41001 ..
XX
XX      Align seg 1/1 to: AAQ41001 from: 1 to: 1749
XX
XX      1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnVal 17
XX      |||||||
XX      107 TTTCAGAGTGGCCAAAGTTCTAGCGCTCTTCCTAGAACCTCTTAGCAGAGT 156
XX      |||||||
XX      17 GlnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnP 34
XX      |||||||
```

```
157 TCAAGTTCTACAGAAATCTTACTACAAACATATGAGATTTGTTCTCTGGCAGC 206
34 rovalThrAlaAspleuIleValLysLysGlnValHisPhePheVal 50
|||||
207 CGGTAAACAGCTGACCTTATGTGAGAAAAAACAAGCATTTTCTTTTGTATA 256
51 AsnAlaSerAspValAspAsnValLysAlaHisIleAsnValSerGlyTl 67
|||||
257 ATATGATCTGATGTCGACAAATGTGAAAGCCCATTTAAATGTGAGCGGAAAT 306
67 eProCysSerValLeuAlaAspValGluAspleuIleGlnGlnI 84
|||||
307 TCCATGACAGTGTCTGTGGCAGACGTGAAGATCTTATTCACAGCAGCA 356
84 IeSerAsnAspThrValSerProAlaSerAlaSerTyrTyrGluGln 100
|||||
357 TTTCACAGACACAGTACGCCGCCGAGCCTCCGATCTCTACTATGACACAG 406
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluAr 117
|||||
407 TATCCTCAGTAATGAATATCTATCTTGATAGATTTATTAAGTGAAG 456
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluPyr 134
|||||
457 GCATCCTGATATGCTTACAAAAATCCACATTCGATTCCTTGAAGAAGT 506
134 YrProLeuThrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
|||||
507 ACCCACTTATGTTTAAAGTTTCTGGAAAAAGAACAAACGCCAAAAAT 556
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
557 GCCATATGATGATTGACTGTGAATCATGCCAGAAATGATGATCTCTCTGC 606
167 aPheCysLeuThrPheIleGlyHis..... 175
|||||
607 TTTCGTTGTGTTCATAGGCGCATATATCAATCTATAGGATTAATAG 656
175 ..... 175
657 GGCATATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATCCCGGTG 706
176 .....AsnArgMetTrpPar 180
707 GTTAATGTGACGGTTATGACTACTACTATGAAAAAAGATGAATGTGGAG 756
180 GlyAsnArgSerPheThrAlaAsnHisCysIleGlyThrAspleuA 197
757 AAAGAACCGTTCTTCTATGCGAACAAATCATGTCATGGAACAGACCTGA 806
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSer 213
807 ATAGAACTTGTCTCCAAACACTGAGTGAGAGAGTGCATGCAGTTC 856
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVa 230
|||||
857 TCATGCTCGGAACCTACTGTGACTTTATCTGAGTCAAGAACACAGAGT 906
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGluIleLysAlaLe 247
907 GAAGCGAGTGCTAGTTCTTGGAGAAATAATCAACCAAGATTAAAGCAT 956
247 YrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
|||||
957 ACATACAGCATGATTCATCTCCAGCATATAGGTTTCATATTCCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
|||||
1007 ACACGAAAGTAAAGCAAAAGCATGAGGAACCTCTCTAGTACCCAGTGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisG 297
|||||
1057 AGCAGTTCTGTCTATTCAGAAAACTAGTAAAAATACAGGTTATACACATG 1106
```

297 IYHISGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrp 313
 1107 GCCATGCTCAGAAACCTTATACCTAGCTCCTGAGAGTGGGAGCATTTGG 1156
 314 IIEYrAspLeuGlyIleIleYrSer.Phe..... 323
 1157 ACTTATGATTGGGACATCAAAATATTGCTTACAAATTGAACCTTGAGATAC 1206
 324ThSerAsnProProval 329
 1207 GGGCACAATACGATTCTGCTGCGGAGCGCTTACATCAAAACCACTGTA 1256
 330 GILYsLeuLeuProLeuSerLeuLys 338
 1257 GAGAACCTTTTGGCCGCTGCTCTAAAA 1283
 seq_name: /SIDSL/9cdata/geneseg/genesegn-emb1/NA2000.DAT:AAF18005
 seq_documentation_block:
 ID AAF18005 standard; DNA; 1400 BP.
 AC AAF18005;
 DT 14-MAR-2001 (first entry)
 DE Lung cancer associated polynucleotide sequence SEQ ID 24.
 KW Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotoxic; antineoplastic; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 OS Homo sapiens.
 XX WO200055180-A2.
 XX 21-SEP-2000.
 XX 08-MAR-2000; 2000WO-US05918.
 XX 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX Ruben SM;
 XX WPI: 2000-587514/55.
 DR P-PSDB; AAB58129.
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 PI Claim 1; Page 507; 1425bp; English.
 PS Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotoxic; antineoplastic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
 alignment_scores:
 Quality: 902.00 Length: 238
 Ratio: 4.983 Gaps: 4
 Percent Similarity: 76.050 Percent Identity: 75.210
 alignment_block:
 US-09-980-881-3 x AAF18005 ..
 Align seg 1/1 to: AAF18005 from: 1 to: 1400
 157 GYLIEHISAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheI 173
 10 GGAATCCATGCGAGANATGATCTCTCTGCTTGTGTTGTTGAT 59
 173 eGLYHIS..... 175
 60 AGGCCATATACATCAATCTATGGGATATAGGCAATATACCAATCTCC 109
 175 175
 110 TGAGGCTTGTGATTTCTATGTTATGCGGTTGTTATGTGATGTTAT 159
 176Asn.ArgMetTrpArgLysAsnArgSerPheT 186
 160 GAACCTACTCATGAGAAAAAGATCGAATGTGAGAAACCGTTCTTCT 209
 186 YrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
 210 ATGCGAACAATCATTTGCTGATCGAAGACAGCTGAACTTTGCTTCC 259
 203 LysHisTrpCysGluGluGlyAlaSerSerSerSerSerGluTrpTrp 219
 260 AACACATGTTGTGAGAGAGTGCATCCAGTTCTCATGCTCGAAGCTTA 309
 219 rCysGlyLeuTrpProGluSerGluProGluValAlaValAlaSerP 236
 310 CTGTGACCTTATCTCTGAGCAACCAAGTGAAGGAGTGGCTGAGTT 359
 236 heLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 252
 360 TCTTGAGAGAAATATATCAACCAAGTTAAAGCATATCAGCATGATCTCA 409
 253 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLy 269
 410 TACTCCAGCATATAGTGTTCATATTCATATACAGAGTAAGTAACCAA 459
 269 sAspHisGluGluLeuSerLeuValAlaSerGluAlaValAlaIleG 286
 460 AGACCATGAGGAACTGCTCTAGTAGCCAGTAGGAGGAGTTCGTCTATG 509
 286 IuLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
 510 AGAAAACTAGTAAAAATACCAAGGTATATACATGCGCATGGCTCGAANA 559
 303 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyI 319
 560 TTATACCTAGCTCTCGAGGAGGAGGAGATGATATGATTTGGGCAT 609
 319 eLYsTYrSer.Phe..... 323
 610 CAAATATTGCTTAACTTGAACCTTGAGATACGGGACATACGAGTTCT 659
 324ThSerAsnProProvalGILYsLeuLeuProLe 335
 660 TGCAGCGGAGCGTTATACATCAAAACCACTGTAGAGAAAGCTTTGGCGCT 709

```

seq_name: /SIDSI/gcsgdata/geneseq/geneseqn-emb1/NAI995.DAT:AAQ90600
seq_documentation_block:
ID      AAQ90600 standard; DNA; 1215 BP.
AC      AAQ90600;
AD
AE
AF
AG      13-MAR-1996 (first entry)
AH
AI
AJ
AK
AL
AM
AN
AO
AP
AQ
AR
AS
AT
AU
AV
AW      Porcine Tyr-His-Met procarboxypeptidase B coding sequence.
AX
AY
AZ
BA      Procarboxypeptidase B; carboxypeptidase B; Pichia; PCPB;
BB      human serum albumin; premating factor alpha; mating factor alpha;
BC      proCBP; ds.
BD
BE
BF
BG
BH
BI
BJ
BK
BL
BM
BN
BO
BP
BQ
BR
BS
BT
BU
BV
BW
BX
BY
BZ
CA      Sus scrofa.
CB
CC
CD
CE
CF
CG
CH
CI
CJ
CK
CL
CM
CN      Key Location/Qualifiers
CO      CDS 1..1215
CP      /*tag= a
CQ      /product= Porcine procarboxypeptidase B.
CR
CS      W09514096-A1.
CT
CU
CV
CW
CX
CY
CZ
DA      26-MAY-1995.
DB
DC
DD
DE
DF
DG
DH
DI
DJ
DK
DL      16-NOV-1994; 94WO-US13142.
DM
DN
DO
DP
DQ
DR      16-NOV-1993; 93US-0153258.
DS
DT
DU
DV
DW
DX
DY
EZ      (EIL ) LILLY & CO ELI.
FA
FB
FC
FD
FE
FF
FG
FH
FI
FJ
FK
FL
FM
FN
FO
FP
FQ
FR
FS
FT
FU
FV
FW
FX
FY
FZ
GA      Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
GB      Sternner JU, Zhang H;
GC
GD
GE
GF
GG
GH
GI
GJ
GK
GL
GM
GN
GO
GP
GQ
GR
GS
GT
GU
GV
GW
GX
GY
GZ
HA      WPI: 1995-200386/26.
HB
HC
HD
HE
HF
HG
HH
HI
HJ
HK
HL
HM
HN
HO
HP
HQ
HR
HS
HT
HU
HV
HW
HX
HY
HZ
IA      P-PSDB: AAR75131.
IB
IC
ID
IE
IF
IG
IH
II
IJ
IK
IL
IM
IN
IO
IP
IQ
IR
IS
IT
IU
IV
IW
IX
IY
IZ
JA      DNA encoding porcine carboxypeptidase B - used for transforming
JB      host cells, partic. Pichia species; for prodn. of the enzyme
JC
JD
JE
JF
JG
JH
JI
JJ
JK
JL
JM
JN
JO
JP
JQ
JR
JS
JT
JU
JV
JW
JX
JY
JZ
KA      Claim 2; Page 20-21; 34pp; English.
KB
KC
KD
KE
KF
KG
KH
KI
KJ
KK
KL
KM
KN
KO
KP
KQ
KR
KS
KT
KU
KV
KW
KX
KY
KZ
LA      The porcine carboxypeptidase B coding sequence can be place in a
LB      bacterial or pref. Pichia yeast expression vector. The expression
LC      vector further comprises the signal peptide of either human serum
LD      albumin (designated pLGd23 - NRRL-B-21029); premating factor alpha
LE      (designated pF4748 - NRRL-B-21028); mating factor alpha (designated
LF      pF4747 - NRRL-B-21032) or the porcine proCBP signal peptide,
LG      (designated pLGd27 - NRRL-B-21027). The method can be used for
LH      producing large amounts of porcine carboxypeptidase B and when
LI      produced in Pichia yeast, the protein does not need solubilisation
LJ      or folding. The produced enzyme is then used for pref. cleaving
LK      basic residues from the carboxy terminus of proteins.
LL
LM
LN
LO
LP
LQ
LR
LS
LT
LU
LV
LW
LX
LY
LZ
MA      Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other:
MB
MC
MD
ME
MF
MG
MH
MI
MJ
MK
ML
MM
MN
MO
MP
MQ
MR
MS
MT
MU
MV
MW
MX
MY
MZ
NA
NB
NC
ND
NE
NF
NG
NH
NI
NJ
NK
NL
NM
NO
NP
NQ
NR
NS
NT
NU
NV
NW
NX
NY
NZ
OA
OB
OC
OD
OE
OF
OG
OH
OI
OJ
OK
OL
OM
ON
OO
OP
OQ
OR
OS
OT
OU
OV
OW
OX
OY
OZ
PA
PB
PC
PD
PE
PF
PG
PH
PI
PJ
PK
PL
PM
PN
PO
PP
PQ
PR
PS
PT
PU
PV
PW
PX
PY
PZ
QA
QB
QC
QD
QE
QF
QG
QH
QI
QJ
QK
QL
QM
QN
QO
QP
QQ
QR
QS
QT
QU
QV
QW
QX
QY
QZ
RA
RB
RC
RD
RE
RF
RG
RH
RI
RJ
RK
RL
RM
RN
RO
RP
RQ
RR
RS
RT
RU
RV
RW
RX
RY
RZ
SA
SB
SC
SD
SE
SF
SG
SH
SI
SJ
SK
SL
SM
SN
SO
SP
SQ
SR
SS
ST
SU
SV
SW
SX
SY
SZ
TA
TB
TC
TD
TE
TF
TG
TH
TI
TJ
TK
TL
TM
TN
TO
TP
TQ
TR
TS
TT
TU
TV
TW
TX
TY
TZ
UA
UB
UC
UD
UE
UF
UG
UH
UI
UJ
UK
UL
UM
UN
UO
UP
UQ
UR
US
UT
UU
UV
UW
UX
UY
UZ
VA
VB
VC
VD
VE
VF
VG
VH
VI
VJ
VK
VL
VM
VN
VO
VP
VQ
VR
VS
VT
VV
VW
VX
VY
VZ
WA
WB
WC
WD
WE
WF
WG
WH
WI
WJ
WK
WL
WM
WN
WO
WP
WQ
WR
WS
WT
WU
WV
WW
WX
WY
WZ
XA
XB
XC
XD
XE
XF
XG
XH
XI
XJ
XK
XL
XM
XN
XO
XP
XQ
XR
XS
XT
XU
XV
XW
XX
XY
XZ
YA
YB
YC
YD
YE
YF
YG
YH
YI
YJ
YK
YL
YM
YN
YO
YP
YQ
YR
YS
YT
YU
YV
YW
YX
YY
YZ
ZA
ZB
ZC
ZD
ZE
ZF
ZG
ZH
ZI
ZJ
ZK
ZL
ZM
ZN
ZO
ZP
ZQ
ZR
ZS
ZT
ZU
ZV
ZW
ZX
ZY
ZZ

```

28 TTCGAAAGCGGAGAGGTCGTCCTCGTCATGCTTGAAGATGAAATGCAAT 77
17 IGIValLeuGIAsnLeuThrThrThrTygLeuIleValLeuTrpGln 34
78 CAGCTTACTGCTAGTGTGGCCAGCACAGGAGATGTCACCTTGGCAAC 127
34 roValThAlaAspLeuIleValLysLysGluValHisIshpheVal 50
128 CAGATTCTGTCACCAAAATCAAACTCACAGTACAGTGTACTTCGCTGTG 177
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyI 67
178 AAACCAACAAGATATTTTGGCTGTGGAAAGACTTGTGGACAGAAAGACT 227
67 eProCySerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
228 ACAATATGAGTACTCTATTAACAACCTGGAGATCTGTGTCAGAGCTCAGT 277
84 IeSerAsnAspThrValSerProArgAlaSerAlaSerTygTygGln 100
278 TTGCACAGCAGAGTC.....CGTACAACGTGCACACAGTATGAGAG 318
101 TyrHisSerLeuAsnGluIleTySerTrpIleGluPheIleThrGluAr 117
319 TACACACAGCGGGAAACGATCGAGCGGTGGACTTAAGCAAAAGTACCAGTGA 368
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluYst 134
369 AAATCCAGACCTCATCTCTCGCACACCCATCGGACATTCATTTTTAGGA 418
134 yTrpLeuTyValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
419 ACAATATATACCTCCCTCAAGGTT...GGCAAACTGGACAAATRAACCT 465
151 AlaIleTrpIleAspPysGlyIleHisAlaArguIleTrpIleSerProAl 167
466 GCCATTTCATGACACTGTGTGTTCCATGCCAGATGATGATTTCCATGC 515
167 apHeCysLeuTrpPheIle.....173
516 ATTTTGGCAATGGTTGTGTGAGAGAGGCTGTTCTCAGTACGATAGATAGA 565
173173
566 GTCAACATGACAGAAATTCCTCAACAAGCTAGACTTTATGTCCTGGCTGTG 615
174GlyHis.....AsnArgMetTrpAr 180
616 CTCATATGTTGATGGTCATACATCAACCTGCAGACCAAGAAGCAAGTGTGAG 665
180 glyAsnAsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAspLeuA 197
666 AAAGACCCCGCTCAACAAGCTGGCAACTCACTGATTTGGCCACACACCCCA 715
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGluAlaSerSerSer 213
716 ACAGAAATTTT...CATGTGGGTGTGTGCACAACATGAGCCTTCAAGAC 762
214 SerCySerGluThrTygCysGlyLeuTyProGluSerGluProGluVa 230
763 CCGTCGGAGTAGACCTTACTGTGATCTGCTGCACAGCTGCAAAAAGAGAC 812
230 IlyAsnAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTr 247
813 CAAGGCCCTGGCGATTTTATACGCAACAACCTGCTCTCATCAAAAGCAAT 862
247 yIleSerMetHisSerTySerGlnHisIleValAlaProTyGlySerTy 263
863 ACCTGACGATCCACATCATCTCAGAGATGATACATCCCTTATTCCTAT 912
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280

```

913 GATTCAACAACCTCCCGAGAACATGCTGAGTGTGAATTAACCTGGCTAAGGC 962
280 uAlaValArgAlaIleGIuLyThrSerLysAsnThrArgTYrThHisG 297
|||||::: ::||| ::::::::::::::1
963 TGCCCTGTAAAGACTT...GCTACACTGTATGGCACCACCAATACACATAGC 1009
297 LYSGLYSerGIUThrLeuTYRLeuAlaProGIYGIYGLYSAspAPTTP 313
||| ||::: ||:::|| | | | |||||:::|||||||
1010 GCCCAGGAGCTACACAACTATCTCTGCTGGGGGCTGTGATGACTGG 1059
314 ILeTYrAspLeuGIYIleLYSTyrSerPheThr 324
||||| ||||| ||||| ||||| ||||| |||||
1060 GCTTATGACCAAGAGATCAATATATCTTCACCC 1092

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.AAT42494

seq_documentation_block:
ID AAT42494 standard; DNA; 1263 BP.
XX
AC AAT42494:
XX
DT 12-FEB-1997 (first entry)
XX
DE Human pancreatic carboxypeptidase B encoding sequence.
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPEPT;
KW mustard-ribonucleotide; antibody directed enzyme prodng therapy;
KW anti-neoplastic; producty; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; Flab?; PelB; leader; human carboxypeptidase B; ss.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 1..1248
FS CDS
FT /*tag= a
FT /product= preproenzyme
FT 1..39
FT /*tag= b
FT /note= "partial enzyme secretion leader sequence
(pre-sequence); residues -108 to -96"
FT sig_peptide
FT 40..324
FT /*tag= c
FT /note= "pro-sequence; residues -95 to -1"
FT mat_peptide
FT 385..1245
FT /*tag= d
FT /note= "mature enzyme; encodes residues +1 to +307"
XX
FN WO9620011-A1.
XX
PD 04-JUL-1996.
XX
PF 21-DEC-1995; 95MO-GH02991.
XX
PR 16-AUG-1995; 95GB-0016810.
XX
PR 23-DEC-1994; 94GB-0026192.
XX
PA (ZENE ) ZENECA LTD.
XX
PI Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
DR WPI: 1996-321650/32.
DR P-PSDB: AA06172.
XX
XX Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
PT anti:neoplastic prodng
XX
PS Reference Example 14; Page 131; 182pp; English.
XX
A two-component system for anti-tumour therapy comprises a targeting

```

[illegible]

```

180 glysasnargserphtyrAlaasnAsnHscysIleGlyThrAspLeuA 197
|||||
696 AAGAGTCCCTCCACCCATCGATGAGTCTAGCTGATGGCACAGACCCCA 745
197 snArgAsnPhelaserLysHsITrPCysGluGluGlyAlaserSer 213
|||||
746 ACAGAAATTTT...CATGCTGGTTGGTGAATAATGGAGCTCTCGAAGC 792
214 SerCysSerGluThrTyrcysGlyLeuTyrrProGluSerGluProGluVa 230
793 CCCTGATGAAGAACTACTGTGACCTGCCGACAGATCGAAGAGAGAGC 842
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleValat 247
843 CAGAGCCCTGGCTGATTCATCCGACAAACTCTCTTCATCAAGGCAAT 892
247 ylleSerMetHisSerTyrrSerGlnHsIleValPheProTyrrSerTy 263
893 ATCTGACAAATCCATCGTACTCCCAATGATGATCTACCTTACTCATAT 942
264 ThrArgSerLysSerLysAsnHisGluGluLeuSerLeuValAlaSer 280
943 GCTTACAAACTCGTGAGAACATGCTGATGAAATGCCCTGGTAAAGC 992
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrrHsG 297
993 TACTGTGAAGAACTT...GCCTCACTGCACGGCACCAAGTACATATMG 1039
297 LysHisGlySerGluThrLeuTyrrLeuAlaProGlyGlyGlyAspPrr 313
1040 GCCCGGAGCTCAACAACTATCTCTCTGCTGGGGGCTCTGACACATCG 1089
314 IleTyrrAspLeuGlyIleLysTyrrSerPheThr 324
1090 GCTTATGACCAAGAAATGATGATATCTTCAAC 1122
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:AAV41795
seq_documentation_block:
ID AAV41795 standard; DNA; 1263 BP.
AC AAV41795;
XX
XX
XX 20-NOV-1998 (first entry)
XX
XX Human pancreatic carboxypeptidase nucleotide sequence.
XX
XX ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
XX prodrg therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1248
XX FT /*tag-a
XX FT /product- "Pancreatic carboxypeptidase B"
XX FT /note- "No start codon given"
XX
XX MO9835988-A1.
XX
XX 20-AUG-1998.
XX
XX 10-FEB-1998; 98MO-GB00415.
XX
XX 29-OCT-1997; 97GB-0022727.
XX 14-FEB-1997; 97GB-0003104.
XX 18-OCT-1997; 97GB-0022003.
XX
XX (ZENE ) ZENBPA (ZENE ) ZENBPA LTD.
XX Edge MD;
XX PI
XX
XX WPI; 1998-467168/40.

```

```

DR P-PSDB; AAW74476.
XX
XX New modified pro-domain of carboxy-peptidase B - enhances expression
PT of co-expressed proteins for production of recombinant
PT carboxy peptidase or its fusions with antibodies, used, e.g. in
PT enzyme prodrg therapy
XX
XX Example 1; Page 52-53; 83pp; English.
XX
XX The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC modified pro-domain of CPB on a separate gene to enhance recombinant
CC expression. This process can be used to produce recombinant CPB in
CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
CC insulin production and protein sequencing, while its fusions with
CC antibodies are useful in antibody-directed enzyme prodrg therapy. The
CC modified pro-domain provide increased yields of recombinant CPB, possibly
CC by protecting the C-terminus against enzymatic degradation or by
CC increasing intracellular trafficking.
XX
XX Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;
XX
XX Alignment_scores:
XX Quality: 637.50 Length: 361
XX Ratio: 2.602 Gaps: 6
XX Percent Similarity: 67.867 Percent Identity: 36.011
XX
XX alignment_block:
XX US-09-980-881-3 x AAV41795 ..
XX
XX Align seg 1/1 to: AAV41795 from: 1 to: 1263
XX
XX 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVa 17
XX |||||
XX 58 TTGTAAGCGAGAAAGCTGTCCGCTTACCTGAAAGTGAATGAAATCAT 107
XX
XX 17 GlnValLeuGlnAsnLeuThrThrTyrrGluIleValLeuTrpGlnP 34
XX |||||
XX 108 TACATTAATCCGCGAGTGGCCAGACGACGCCAGATTGACTTGTGAAGC 157
XX
XX 34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
XX |||||
XX 158 CAGATTCTGCACAAATCAAACTCAAGTACAGTTCAGTTCCTCGGTT 207
XX
XX 51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyI 67
XX |||||
XX 208 AAAGCAGAAAGTACTGCTGCTGAGAAATGTCTTAAGCAGATGAACT 257
XX
XX 67 eProcysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
XX |||||
XX 258 ACAATACAGAGTACTGATTAAGCAACCTGAGAAATGTGTGAGGCTCAGT 307
XX
XX 84 leSerAsnAspThrValSerProArgAlaSerAlaSerTyrrTyrrGln 100
XX |||||
XX 308 TTGATAGCCGGGTT.....CGTGCACAGACACAGTTATGAGAGAG 348
XX
XX 101 TyrHisSerLeuAsnGluIleTyrrSerTrpIleGluPheIleThrGlu 117
XX |||||
XX 349 TACAACAAGTGGAAAGCATGAGAGCTTGAGCTCAACAAGTGGCAGCTGA 398
XX
XX 117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
XX |||||
XX 399 GAATCCAGCCCTCATCTCTCGAGTGTATCGGAAACCATTTGAGGAGAC 448
XX
XX 134 yrrProLeuTyrrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
XX |||||
XX 449 GCGGTATTTTACTCTTAAGGTT...GCCAAGCTGACAAATAAGCCT 495
XX
XX 151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
XX |||||
XX 496 GCCATTTCATGAGACTGTGTTTCATGATGCCAGAGAGATTCCTCTCTGC 545
XX
XX 167 apheCysLeuThrPheIle.....
XX

```



```

335 TTTGATAGCCCGGTT.....CGTGCAACAGGACACACTTATGAGAAAG 375
101 TTTTATSerLeuAaNGluIleuTyrSerTrpIleGluPheIleThrGluAT 117
376 TACAAACAAGTGGGAAACGATAGAGCGGTGGACATCAACAATCCCACTGA 425
117 gHISProAspMetLeuThrIleuHISileHISileGlySerSerPheGluLysT 134
426 GAATCCAGCCCTCATCTCTCGCAGGTATTATCGAACCCATTGAGGAGAC 475
134 YrProLeuTyrValLeuLysValSerGlyLysGluInThrAlaLysAsn 150
476 GCGCTATTATACCTCTCGAAGCTT...GCCAAAGCTGACAAATAAGCCT 522
151 AlAlIleTrpIleAspCysGlyIleHISAlaArgGluTrpIleSerProAl 167
523 GCCATTTTCATGAGACTGTGGTTTCATGCCAGAGCTGATTTCTCTCTG 572
167 aPheCysLeuTrpPheIle..... 173
573 ATTCTGCAGGTGGTTGTAAAGAGAGCGCTTCTGACTATGAGAGTGAAG 622
173 ..... 173
623 TCCAAAGTGAAGAGCTTCTGACAAAGTATGACTTTATGCTCCGCTGTG 672
174 .....GlyHIS.....AsnArgMetTrpArg 180
673 CTCCAATATTGATGGCTATACATCTACACCTGACCAAGAGCCGATTTGGAG 722
180 gHISAsnArgSerPheTyrAlaAsnHIScysIleGlyTrpIleAspLeuA 197
723 AAGAGCTGCTCCACCCATCTGAGATCTGATGCTGATGGACAGACCCCA 772
197 snArgAsnPheAlaSerLysHISTrpCysGluGluLysLysSerSer 213
773 ACAGAAATTTT...GATGCTGGTGGTGAATGGAATGGACCTCTCGAAAC 819
214 SerCysSerGluTrpTyrCysGlyLeuTyrProGluSerGluProGluVa 230
820 CCCTGTGAGAAACTTACTGACCTGCCGACAGTCTGAAAGAGAC 869
230 lLysAlaValAlaSerPheLeuArgArgAsnIleAsnGluIleLysAlaT 247
870 CAAGGCCCGGTGATTTATCCGACAAACAATCTCTTCCATCAAGGCAAT 919
247 YrIleSerMetHISerTyrSerGlnHISileValPheProTyrSerTyr 263
920 AACTGACAAATCCACTGCTACTCCAAATGATGATGATCACTTACATCAT 969
264 ThrArgSerLysSerLysAspHISGluGluLeuSerLeuValAlaSerG 280
970 GCTTTCAAACCTCGGTGAGAACATCTGAGTGAATGCCCTGGCTAAAGC 1019
280 uAlaValAlaArgAlaIleGluLysThrSerLysAsnThrArgTyrTrpHIS 297
1020 TACTGTGAAGAAACTT...GCCCTACTGACGGGACCAAGTACATATG 1066
297 lYHISGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrp 313
1067 GCCCGGAGGCTAACAAATCTATCTCTGCTGCGGCGCTTGCACAGCTGG 1116
314 lIleTyrAspLeuGlyIleLysTyrSerPheTrp 324
1117 GCTTATGACCAAGAAATCAAGATATCTTTCAC 1149
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAD03837
seq_documentation_block:
ID AAD03837 standard; cdna; 1311 bp.
XX
AC AAD03837;

```

```

XX
DT 19-JUN-2001 (first entry)
XX
DE Human carboxypeptidase-related protein #5 CDNA.
XX
KM Human; carboxypeptidase-related protein; protease; gene therapy;
XX drug screening; mental; biological disorder; medical disorder; ss.
OS Homo sapiens.
XX
FH Key location/qualifiers
FT CDS 1..1311
FT /*tag= a
FT /product= "Human carboxypeptidase-related protein #5"
FT /note= "CDS does not include stop codon"
XX
PN WO200123588-A1.
XX
PD 05-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US26876.
XX
PR 29-SEP-1999; 99US-0156685.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI: 2001-266171/27.
XX DR P-PSDB; AAE00502.
XX
PT New isolated human carboxypeptidase polynucleotide useful for
PT generating antibodies, as reagents in diagnostic assays and for
PT screening for compounds useful for treating mental, biological or
PT medical diseases
XX
PS Claim 1; Page 30; 36pp; English.
XX
CC The present sequence is a cDNA encoding novel human protein (NHP),
CC carboxypeptidase-related protein. The carboxypeptidase-related
CC proteins share structural similarity with animal proteases, and
CC especially carboxypeptidase B or carboxypeptidase A. The
CC carboxypeptidase-related protein cDNA is useful for the detection of
CC mutant human carboxypeptidases, or inappropriately expressed human
CC carboxypeptidases for the diagnosis of disease, and also as a
CC therapeutic. It is also useful in drug screening, for generation of
CC antibodies, as reagents in diagnostic assays, for the identification of
CC reagents in assays for screening for compounds related to human carboxypeptidases, and as
CC treating mental, biological or medical disorders and diseases. Nucleotide
CC constructs encoding functional NHPs can also be used in gene therapy.
XX
SQ Sequence 1311 bp; 391 A; 278 C; 284 G; 358 T; 0 other;

alignment_scores:
Quality: 619.50 Length: 391
Ratio: 2.560 Gaps: 9
Percent Similarity: 61.893 Percent Identity: 34.783

alignment_block:
US-09-980-881-3 x AAD03837 ..

Align seg 1/1 to: AAD03837 from: 1 to: 1311

1 PheGlnSerGlyGlnValLeuAlaLeuProArgTrpSerArgGlnVa 17
: : : : : : : : : : : : : : : : : : : : : : : : : :
109 TATGCTGCTGATAAAGTATAGATTATTTCCAAACAGAGAGGAAAGC 158
: : : : : : : : : : : : : : : : : : : : : : : : : :
17 lGlnValLeuGlnAsnLeuThrTrpTrpTyrGluIleValLeuTrpGlnP 34
: : : : : : : : : : : : : : : : : : : : : : : : : :
159 ATATGCACTGAAGAAATAATCTTATCAACTTAAGCTGAGACTGTGGAGC 208

```


alignment_block:
US-09-980-881-3 x AAD090601 ..

Align seg 1/1 to: AAD090601 from: 1 to: 921

```

98 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIle 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 TATGAGAGTACAAACAACTGGGAACGATCGAGCTTGAGCTTAAGCAAC 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 ethrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerP 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 CACCAAGTGAATAATCCAGACCTCATCTCTGCACAGCAATCGGAATCAT 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 heGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThr 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 TTTTAGCAAAACAATATATACCTCTCAAGTTT...GCCAAACCTGGACCA 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 AAlaLysAsnAlaIleTrpIleAspGlyIleHisAlaArgGluTrpIle 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 AATAAGCCTGCCATTTTCATGGAGCTGGTTCCATGCCAGAGAAATGAT 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 eSerProAlaPheCysLeuTrpPheIle..... 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 TTCCCATGATTTTTCAGAGTGTGTGTGAGAGAGGCTGTCTCACCCTATG 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 ..... 173
263 GATATAGAGTGCATCATGACAGAAATTCCTCAACAACTAGACTTTTATGTC 312
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 .....GlyHis.....AsnArg 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 TTGCTGTGCTCATATATGATGGCTACATCTACACCTGGACCAAGACCG 362
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177 gMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyT 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 AATGTGGAGAAAGACCGCTCTACCAATGCTGGAGTACCTGATGGCA 412
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 hArgPLeuAsnArgAsnPheAlaSerLysHisIleTrpCysGluGluLysAla 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413 CAGACCCCAACAGAAATTTT...GATGCTGGTGGTGCACAACTGACACC 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 SerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerG 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 TCTACAGACCCCTGGATGAGACTTACGTGTGATGCTGCGAGACTCTGCA 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227 uProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnI 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 AAAAGAGCAAGGCCCTGGCTGATTTATAGCCAAACCTCTCTCCA 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 LeLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 TCMAAGCATACCTGACGATCCACTCATCTACAGATGATACTACTACCT 609
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuVal 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 TATTCCTGTGATTTACAAACCTCCCGAGAAACAATGCTGAGTTGAATTAAC 659
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
277 lAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgT 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 GGCTAAGGCTGCCGTGAAGAACTT...GCTACACTGTATGACCAACAGT 706
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
294 YrThrHisGlyHisIleSerGluThrLeuTyrLysAlaProGlyGly 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 ACAATACGCGCCAGAGACTACAAACAATCTATCTCTGCTGGGGGCTCT 756
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 AspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 324
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 GATGACTGGGCTTATGACCAAGAAATCAATATCTCTTCAACC 798
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /STDS1/gcgdata/geneseq/geneseg-emb1/NA2001A.DAT: AAD03839

seq_documentation_block:
ID AAD03839 standard; DNA: 2128 BP.

```

XX XX AAD03839;
AC AC
XX XX
DE DE 19-JUN-2001 (first entry)
XX XX
DE DE Human carboxypeptidase-related protein DNA with flanking 5' and 3' ends.
XX XX
KW KW Human; carboxypeptidase-related protein; protease; gene therapy;
XX XX drug screening; mental; biological disorder; medical disorder; ds.
OS OS Homo sapiens.
XX XX
XX XX WO200123588-A1.
XX XX
XX XX 05-APR-2001.
XX XX
XX XX 29-SEP-2000; 2000WO-US26876.
XX XX
XX XX 29-SEP-1999; 99US-0156685.
XX XX
XX XX (LEXI-) LEXICON GENETICS INC.
XX XX
XX XX Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX XX WPI; 2001-266171/27.
XX XX
XX XX New isolated human carboxypeptidase polynucleotide useful for
XX XX generating antibodies, as reagents in diagnostic assays and for
XX XX screening for compounds useful for treating mental, biological or
XX XX medical diseases -
XX XX
XX XX Disclosure; Page 33; 36pp; English.
XX XX
XX XX The present sequence is a novel human protein (NHP),
XX XX carboxypeptidase-related protein DNA with 5' and 3' flanking ends. The
XX XX carboxypeptidase-related proteins share structural similarity with
XX XX animal proteases, and especially carboxypeptidase B or
XX XX carboxypeptidase A. The carboxypeptidase-related protein cDNA is useful
XX XX for the detection of mutant human carboxypeptidases, or inappropriately
XX XX expressed human carboxypeptidases for the diagnosis of disease, and also
XX XX as a therapeutic. It is also useful in drug screening, for generation of
XX XX antibodies, as reagents in diagnostic assays, for the identification of
XX XX other cellular gene products related to human carboxypeptidases, and as
XX XX reagents in assays for screening for compounds that are useful for
XX XX treating mental, biological or medical disorders and diseases. Nucleotide
XX XX constructs encoding functional NHPs can also be used in gene therapy.
XX XX
XX XX Sequence 2128 BP; 640 A; 461 C; 436 G; 591 T; 0 other;

```

alignment_scores:
Quality: 571.50 Length: 441
Ratio: 2.342 Gaps: 13
Percent Similarity: 55.329 Percent Identity: 31.746

alignment_block:
US-09-980-881-3 x AAD03839 ..

Align seg 1/1 to: AAD03839 from: 1 to: 2128

```

3 SerGlyGlnValLeuAlaAlaLeuProArgThrSerArgLysValGlnVal 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
461 AGTGAATAAGTAAAGATTTATTTCCAAAACAGAGAGCAATGTC 510
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
19 LeuGlnAsnLeuThr.ThrThyGluIleVal..... 30
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511 ACTGAAGAAATATCTATCACTTAAGTTGGTTCTCGCACTACAGTG 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31 ...LeuTrpGlnProValThrAlaAspLeuIleValLys.....LysLys 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 GACCTGTGGCAGCCACAGATATCTCTATGTATACAGAGGAAACAGCTTAC 610
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
44 sGlnValHisPhePheValAsnAlaSerAspValAlaAspAsnValLysAla 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


This Page Blank (uspto)

OM of: US-09-980-881-3 to: Issued_Patents_NA:* out_format : pfs

Date: Sep 18, 2002 6:55 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frim+g2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US09980881/rn1.16092002.140129.10653/app.query.fasta.1.1126
-DB=Issued_Patents_NA -OPT=fastap -SUFFIX=rl -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPT=0.000 -LOOEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -FGAPOP=10.000 -FGAPEXT=0.500
-DELTA=6.000 -DELTA=7.000 -START=1 -MATRIX=10sum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 -CGN1_1.62 -NCPU=6 -ICPU=3 -LONGLOC
-DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-3
Query length: 338
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 90.520000

score_list:

Sequence	Strid Orig	ZScore	EScore	Len	Documentation	..
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1 +	1732.00	3887.62	7.5e-209	1272	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-649-591B-2 +	1732.00	3883.43	1.3e-208	1749	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-277-540-2 +	1732.00	3883.43	1.3e-208	1749	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-430-787A-2 +	1732.00	3883.43	1.3e-208	1749	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-696-139-1 +	646.50	1426.09	9.6e-72	1215	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-56 +	637.50	1405.17	1.4e-70	1263	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-71 +	637.50	1404.95	1.4e-70	1284	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-171-945-124 +	619.50	1357.31	6.5e-68	2154	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-782-760-5 +	610.00	1346.87	2.5e-67	927	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:PCITUS-US96-00995-5 +	610.00	1346.87	2.5e-67	927	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-696-139-3 +	582.00	1283.44	8.4e-64	921	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-67 +	567.00	1248.35	7.6e-62	999	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-64 +	567.00	1247.66	8.3e-62	1053	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-74 +	560.00	1231.70	6.4e-61	1059	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-77 +	559.00	1229.43	8.6e-61	1059	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-171-945-112 +	549.00	1199.26	4.1e-59	1870	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-09-395-936-3 +	516.50	1130.84	2.7e-55	1251	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-09-395-936-1 +	516.50	1130.84	2.7e-55	1251	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-640-906-1 +	491.50	1074.07	3.9e-52	1257	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-395-936-1 +	491.50	1074.07	3.9e-52	1257	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-331-709-2 +	210.50	446.65	3.4e-17	591	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-996-416-125 -	169.50	355.47	4.1e-12	515	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-484-105-15 +	116.50	214.42	0.0003	2504	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-484-106-15 +	116.50	214.42	0.0003	2504	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-706-216-1 +	102.00	180.44	0.0231	2719	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-782-760-2 +	97.00	198.81	0.0022	285	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:PCITUS-US96-00995-2 +	97.00	198.81	0.0022	285	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-915-203-1 +	94.00	157.90	0.4164	3797	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-277-887-1 +	94.00	157.90	0.4164	3797	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-789-449-1 +	94.00	157.90	0.4164	3797	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-609-049A-29 +	91.00	146.74	1.74	5285	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-170-996-29 +	91.00	146.74	1.74	5285	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-609-049A-27 +	91.00	143.36	2.69	6831	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-170-996-27 +	91.00	143.36	2.69	6831	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-073-237-1 +	90.50	163.75	0.1367	1333	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-689-974-2 -	88.50	157.01	0.4669	1576	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-09-058-376-2 -	88.50	157.01	0.4669	1576	1	
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-233-989-1 +	88.50	151.26	0.9765	2439	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-452-262-1 +	88.50	151.23	0.9792	2443	1	
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-734-550-1 +	88.50	151.23	0.9792	2443	1	

/cgn2_6/ptodata/1/ina/5B.COMB.seq:PCITUS-US96-07528-1 + 88.50 151.23 0.9792 24
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-412-102-5 + 87.50 155.41 0.5733 1498
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-217-787-5 + 87.50 155.41 0.5733 1498
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-254-358-1 - 86.00 137.00 6.08 4680
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-475-391-1 - 86.00 137.00 6.08 4680

seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1

seq_documentation_block:

Sequence 1, Application US/08869057
Patent No. 5985562
GENERAL INFORMATION:
APPLICANT: Morser, Michael J
TITLE OF INVENTION: Method of Detecting Thrombotic Disease
TITLE OF INVENTION: Risk
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Herlex Biosciences Legal Department
STREET: 15049 San Pablo Avenue
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94804-0099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,057
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Washien, Wendy L.
REGISTRATION NUMBER: 36,301
REFERENCE/DOCKET NUMBER: 51509AUSM1
TELEPHONE: 510-262-5411
TELEFAX: 510-262-7095
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1272 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PUBLICATION INFORMATION:
AUTHORS: Eaton, Dan L.
AUTHORS: Malloy, Beth E.
AUTHORS: Tsai, Siac P
AUTHORS: Henzel, William
AUTHORS: Dryden, Dennis
TITLE: Isolation, Molecular Cloning, and Partial
TITLE: Characterization of a No. 5985562 Carboxypeptidase B
JOURNAL: J. Biol. Chem.
VOLUME: 266
ISSUE: 32
PAGES: 21833-21838
DATE: No. 5985562 15-1991
US-08-869-057-1

alignment_scores:
Quality: 1732.00 Length: 393
Ratio: 5.124 Gaps: 3
Percent Similarity: 86.005 Percent Identity: 85.751

alignment_block:
US-09-980-881-3 x US-08-869-057-1 ..
Align seg 1/1 to: US-08-869-057-1 from: 1 to: 1272

```

1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgLys 17
   |||
67 TTTTCAGAGTGGCCAGATGTCAGCTGCTTCTTAGAACCTTAGGCAAGT 116
   |||
17 LglnValLeuGlnAsnLeuThrThrThrArgLysValLeuLeuTrpGln 34
   |||
117 TCAGAGTTCAGAAATCTTACATACATATGAGATTGTTCTCTGCGACG 166
   |||
34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
   |||
167 CGGTACAGCTGACCTTATTGTGACAAAAAACATCCATTTTTCATGTA 216
   |||
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly 67
   |||
217 AATGCATCTGATGTCGACAAATGCAAGCCCATTTAAATGTGAGCGCAAT 266
   |||
67 eProcysSerValLeuLeuAlaAspValGlnAspLeuIleGlnGln 84
   |||
267 TCCATGCACTGCTTGTGTCGACAGCTGGAAAGTCTTATTCAACACGACA 316
   |||
84 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGln 100
   |||
317 TTTCCAAAGACACAGTCAAGCCCCGAGCCTCCGCACTGACTATGACAG 366
   |||
101 TyrHisSerLeuAsnGlnIleTyrSerTrpIleGluPheIleThrGln 117
   |||
367 TATCACTCTCAATGAATGAAATCTATCTTGATAGAAATTTATTAACGAGAG 416
   |||
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
   |||
417 GCATCTGATGATGCTTACAAAAATCCACATTTGATCTTATTTGAGAACT 466
   |||
134 YrProLeuTyrValLeuLysValSerGlyLysGlnGlnThrAlaLysAsn 150
   |||
467 ACCCACTCTATGTTTAAAGTTTCTGGAAAGAACCAACACCAAAAT 516
   |||
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
   |||
517 GCCATATGATGATGACGTGGAAATCCATGCGACAGAAATGATCTCTCC 566
   |||
167 aPheCysLeuTrpPheIleGlyHis..... 175
   |||
567 TTTTCGCTGTGTTCATATGCGCATATATCAATTCATGCGATAATAG 616
   |||
175 ..... 175
   |||
617 GGCATATATACCAATCTCCTGAGGCTTGTGATTTCTATGATGCCGAG 666
   |||
176 .....AsnArgMetTrpArg 180
   |||
667 GTTAAATGTCGACGTTATGACTACTCATGAGAAAAAATCGAATGTGAG 716
   |||
180 gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 197
   |||
717 AAAGAACCGTCTTCTTATGCGAACATCATGATCGAAGACAGACGTGA 766
   |||
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 213
   |||
767 ATAGGAATTTGCTTCCAAACACTGTGTGAGGAGGTGCATCCAGTCC 816
   |||
214 SerCysSerGluThrTyrCysGlyLeuTyrProGlnSerGluProGln 230
   |||
817 TCATCTCTGGAACCTACTGTGAGATTATCTCTGATGAGAACCGAAGT 866
   |||
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAla 247
   |||
867 GAAGGCACTGGCTAGTTCTTGAGAAAGAAATATCAACAGATTAAAGCAT 916
   |||
247 YrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
   |||
917 ACATCAGCATCATTCATATCTCCAGCATATAGTTCATATTCATCTAT 966

```

```

264 ThrArgSerLysSerLysAspHisGlnGluLeuSerLeuValAlaSer 280
   |||
967 ACACGAAAGTAAAGCAAAACCATGAGGAACTGTCTCTGTAGCCGCTGA 1016
   |||
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisG 297
   |||
1017 AGCAGTTCGTGCTATGAGAAAACTAGTAAAAATACCAAGTATACAGT 1066
   |||
297 LysIleGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspTrp 313
   |||
1067 GCCATGCTCAGAAACCTTATACCTAGCTCTGAGAGTGGGACGATTTGG 1116
   |||
314 IleTyrAspLeuGlyIleLysTyrSer.Phe..... 323
   |||
1117 ATCTATGATTGGCCATCAAAATTTGTTTACAATTGAACCTCGAGATAC 1166
   |||
324 .....ThrSerAsnProProVal 329
   |||
1167 GGGCACTACGGAATCTTCTGCGGAGCGCTTACATCAAAACCACTGTA 1216
   |||
330 GlnLysLeuLeuProLeuSerLeuLys 338
   |||
1217 GAGAACTTTTCCGCTGCTCTTAAAA 1243

```

seq_name: /cgn2/6/plodata/1/lna/5A_COMB.seq:US-07-649-591B-2

seq_documentation_block:

Sequence 2, Application US/07649591B

Patent No. 5206161

GENERAL INFORMATION:

APPLICANT: Dennis Drayna and Daniel Eaton

TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patlin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/649,591B

FILING DATE: 19910201

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Hasak, Janet E.

REGISTRATION NUMBER: 28,616

REFERENCE/DOCKET NUMBER: 689

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/266-1896

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1749 bases

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: hybridization probe

LOCATION: 133 to 178

IDENTIFICATION METHOD:

OTHER INFORMATION:

FEATURE:

NAME/KEY: potential clip site


```
LOCATION: 380 to 382
IDENTIFICATION METHOD:
OTHER INFORMATION:
FEATURE:
NAME/KEY: signal sequence
LOCATION: 41 to 106
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-07-649-591B-2
```

```
alignment_scores:
    Quality: 1732.00    Length: 393
    Ratio: 5.124        Gaps: 3
    Percent Similarity: 86.005    Percent Identity: 85.751
```

```
alignment_block:
US-09-980-881-3 x US-07-649-591B-2 ..
```

```
Align seg 1/1 to: US-07-649-591B-2 from: 1 to: 1749
```

```
1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVal 17
|||||
107 TTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTGAACTCTAGCAAGT 156
17 GlnValLeuGlnAsnLeuThrThrThrGluIleValLeuTrpGlnP 34
|||||
157 TCAAGTTCTACAGAACTTCTACTACACATATGAGATTGTCTCTGCGAGC 206
34 rovalThrAlaAspLeuIleValLysLysGlnValHisPheVal 50
|||||
207 CGGTAACAGCTGACCTTATGTGAAAGAAAAACAAGTCCATTCTTTGTA 256
51 AsnAlaSerAspValAlaAsnValLysAlaHisLeuAsnValSerGly 67
|||||
257 AATGCATCTGATGCGACAAATGTAAGCCCATTTAAATGCGAGCGAAT 306
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 84
|||||
307 TCCATGCGAGTGTCTTCTGCGACGCGTAAGATCTTATTCACACAGAGA 356
84 leSerAsnAspThrValSerProArgAlaSerAlaSerTyrrGluGln 100
|||||
357 TTTCACAGACAGACAGTCAGCCCGCGAGCTCTCGCATGCTATGAAACAG 406
101 TyrHisSerLeuAsnGluIleTyrrSerTrpIleGluPheIleThrGlu 117
|||||
407 TATCAGTCACTAAATGAATCTATCTTGGATAGAAATTTTAACTGAGAG 456
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
|||||
457 GCATCCGATATGCTTACAAAACCAACATGATGATCTCATTTGAGAGT 506
134 yrrProLeuTyrrValLeuLysValSerGlyLysGluIleThrAlaLys 150
|||||
507 ACCCAGCTGATGTTTAAAGGTTCTTGAAAAAGCAAAACGCCAAAAT 556
151 AlaIleTrpIleAspCysGlyIleHisAlaArgLutTrpIleSerProAl 167
|||||
557 GCCATATGATGATGACTGTGGAATCCATGCCAGAAATGATCTCTCTGCC 606
167 aPheCysLeuTrpPheIleGlyHis..... 175
607 TTTCGCTGTGGTTTCATAGCCATATAACATCATCTATGAGATAATAG 656
175 ..... 175
657 GGCATATPACCAATCTCCTGAGGCTTGATTTCTATGTTATGCCGGTG 706
176 .....AsnArgMetTrpArg 180
707 GTTAATGTGGACGGTTATGACTACTCATGAAAAAGAAATGCAATGTGAG 756
```

```
180 glysAsnArgSerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeu 197
|||||
757 AAGAACCGTCTTCTATGCGAACAATCATGATGCGAACAAGACAGCTGA 806
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 213
|||||
807 ATAGAACTTGTCTTCCAAACACTGGTGTGAGAAAGGCGCATCCAGTTCC 856
214 SerCysSerGluThrTyrrCysGlyLeuTyrrProGluSerGluProGlu 230
|||||
857 TCATGCTCGGAACCTACTGCTGAGCTTATCTGATGACAGAACAGAAAGT 906
230 llysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAla 247
|||||
907 GAAGCAGTGGCTAGTTCTTGTGAGAGAAATATCAACCGATTTAAAGCAT 956
247 yrrIleSerMetHisSerTyrrSerGlnHisIleValPheProTyrrSer 263
|||||
957 ACATCAGCATCATCTACTACTCCAGCATATAGTGTTCATATTCCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
|||||
1007 ACAGAAAGTAAAGCAAAAGACCATGAGGAACTGTCTAGTAGCAGTGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrrHisG 297
|||||
1057 AGCAGTTCGTCTATTTGAGAAACTAGTAAATAACAGTATACACATG 1106
297 lYHisGlySerGluThrLeuTyrrLeuAlaProGlyGlyGlyAspAsp 313
|||||
1107 GCCATGCTCAGAAACCTTATACCTAGCTCTCGGAGGCGGAGCATGG 1156
314 lIleTyrrAspLeuGlyIleLysTyrrSer.Phe..... 323
|||||
1157 ATCTATGATTTGGCGATCAATATATTCGTTTACATTTGAACTCGAGATAC 1206
324 .....ThrSerAsnProProVal 329
1207 GGGCACAATACGAAATCTTGTCTGCGGAGCGTTACATCAAAACCACTGTA 1256
330 GluLysLeuLeuProLeuSerLeuLys 338
|||||
1257 GAGAAAGCTTTGCCGCTGTCTTAAAA 1283
seq_name: /cgn2_6/prodata/1/lna/5A_COMB.seq:US-08-277-540-2
seq documentation block:
; Sequence 2, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
```

```

APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-277-540-2

```

```

Alignment scores:
Quality: 1732.00      Length: 393
Ratio: 5.124          Gaps: 3
Percent Similarity: 86.005      Percent Identity: 85.751

```

```

Alignment block:
US-09-980-881-3 x US-08-277-540-2

```

```

Align seg 1/1 to: US-08-277-540-2 from: 1 to: 1749

```

```

1 PheGInserGlyGlnValLeuAlaLeuProAlGlyThSerArgGlnVa 17
107 TTTCAGAGGCGCAAGTCTAGCTGCTCTCTCTAGAACCTTAGCGCAGT 156
157 TCAGAGTCTACAGAACTTACTACACAAATAGATGTGCTCTGCGCAGC 206
17 IGInValLeuGlnAsnLeuThrThrThyGlnIleValLeuTrpGlnP 34
157 TCAGAGTCTACAGAACTTACTACACAAATAGATGTGCTCTGCGCAGC 206
34 rovalThAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
207 CGGTACAGCGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTGTGA 256
51 AsnAlaSerAspValAlaAspAsnValLysAlaHisLeuAsnValSerGlyTl 67
257 AATGATCTGATGTGACAAATGTGAAGCCATTAAATGTGAGCGGAGT 306
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
307 TCCATGACAGTGTCTGTGGCAGAGCTGGAAGATCTTATTCACAGCAGA 356
84 IeSerAsnAspThrValSerProAlGAlaSerAlaSerTyTyGln 100
357 TTTCACAGCAGACAGTCCAGCCCGGAGCTCCGCAATCTACTATCAACAG 406
101 TyrHisSerLeuAsnGlnIleTySerTrpIleGluPheIleThGluAr 117
407 TATCCTCTACTAAATGAATCTATCTTGATAGAAATTTAACTGAGAG 456
117 GHISProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysT 134
457 GCATCTCTATGTCTTACAAAAATCCACATGATGATTCCTCATTTGAGAGT 506
134 yTrProLeuTyValLeuLysValSerGlyLysGluGlnThAlaLysAsn 150
507 ACCCAGCTCTATGTTTAAAGGTTTCTGAAAAAAGCAAAAGCCAAAAAT 556
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
557 GCCATATGAGATTGACTGTGGAATCCATGCCAGAAATGATCTCTCTGCC 606
167 aPhecysLeuTrpPheIleGlyHis..... 175

```

```

|||||
607 TTTCCTCTGTGCTCATAGCCATATTAACCAATTCATGGGATTAATAG 656
175 ..... 175
657 GCCAATATACCAATCTCCTAGGCGTTGGATTTCTATGTATGCCGGTG 706
176 .....AsnArgMetTrpAr 180
707 GTTAATGTGAGGTTATGACTACTACATGCAAAAAAGAAATGCAATGTGAG 756
180 gLysAsnArGserPheTyAlaAsnAsnHisCysIleGlyThraPleuA 197
757 AAAGAACCGTCTTCTTATGCGAACAAATCATTTGATCGAACAGACCTGA 806
197 snArgAsnPheAlaSerLysHisIleTyPcysGluGlnGlyAlaSerSer 213
807 ATAGGACTTCTCTCCAAACAGCTGTGTGAGGAAGGTCCATCCAGTTCC 856
214 SerCysSerGluThrTyrcysGlyLeuTyTrpGluSerGluProGluVa 230
857 TCATGCTCGAAGAACTACTGTGACTTATCCTGATCGAAGCAGAACT 906
230 lLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaT 247
907 GAAGGACAGTGTCTTCTGAGAAAGAAATATCAACCAATTAAGCAT 956
247 yTrIleSerMetHisSerTySerGlnHisIleValPheProTySerTy 263
957 ACATGACATGATTCATCATCTCCAGCATATATAGTGTCCATATTCCTAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
1007 ACACGAAGTAAAGCAAGAACATGAGAGACTGTCTCTAGTAGCAGAGGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyThrHisG 297
1057 AGCAGTCTGCTATGTGAGAAACTAGTAAATAACAGGTATACACATG 1106
297 lHisGlySerGluThrLeuTyTrpLeuAlaProGlyGlyLysAspTrp 313
1107 GCCATGCGCTCAGAAACCTTATACCTAGCTCCTGAGAGTGGGAGCATTCG 1156
314 lLeTyAspLeuGlyIleLysTySer.Phe..... 323
1157 ATCTATGATTTGGGCATCAAAATATCTGTACAAATGAACTTCGAGATAC 1206
324 .....ThrSerAsnProProVal 329
1207 GGGCAGATACGGAATTTGCTGCCGAGACGTTACATCAAAACCCACGTGA 1256
330 GluLysLeuLeuProLeuSerLeuLys 338
1257 GAGAAAGCTTTTGGCGGTGTCTTAAAA 1283

seq_name: /cgn2_6/prodata/1/lna/5A_COMB.seq:US-08-430-787A-2

seq_documentation_block:
; Sequence 2, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689DLC1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEX: 415/952-9881
FAX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-430-787A-2

alignment_scores:
Quality: 1732.00 Length: 393
Ratio: 5.124 Gaps: 3
Percent Similarity: 86.005 Percent Identity: 85.751

alignment_block:
US-09-980-881-3 x US-08-430-787A-2 ..

Align seg 1/1 to: US-08-430-787A-2 from: 1 to: 1749

```
1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVal 17
|||||
107 TTTCAGAGTGGCCAGATTCTAGCTCTCTCTAGACCTCTAGCAGT 156
17 GlnValLeuGlnAsnLeuThrThrTYRGIUlleValLeuTrrGlnP 34
|||||
157 TCAGTTCTCAGAAATCTTACTACAAACATATGAGATGTTCTGCGCAG 206
34 roValThrAlaAspLeuIleValIleLysLysGlnValAlaHisPhePheVal 50
|||||
207 CGGTAAACGCTGACCTTATGTGTGAGAAAAAACAGAGTCCATTTTTTGTGA 256
51 AsnAlaSerAspValAspAsnValIleAlaHisLeuAsnValSerGlyI 67
|||||
257 AATGCATCTGATGTGCACAAATGTGAAGCCCATTTAAATGTGACCGCAAT 306
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnI 84
|||||
307 TCATATGAGTGTCTGTCTGCGACGCTGGAAGATCTTATTCACACAGAGA 356
84 IeSerAsnAspThrValSerProArgAlaSerAlaSerTYRGIUlleVal 100
|||||
357 TTTCACACGACACAGTACGCCCGGACCTCCCGCATCTGATATGAAACAG 406
101 TYRHisSerLeuAsnGluIleTYRSerTrrIleGluPheIleThrGluAr 117
|||||
407 TATCACTACCTAAATGAATCTATTTTGGATGAAATTTTAACTGACGAG 456
```

```
117 GHISProAspMetLeuThrIleHisIleGlySerSerPheGluIysT 134
|||||
457 GCATCCGATATGCTTACAAAAATCCACATGGATGCTCATTTAGAGAGT 506
134 YRPProLeuTYRValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
|||||
507 ACCCACTCTATGTTTAAAGGTTCTGGAAGAAACAAACAGCCAAAT 556
151 AlaIleTrrPheAspCysGlyIleHisAlaArgGluTrrPheSerProAl 167
|||||
557 GCCATATGATATGACTGAGGAATCCATGCCAGAAATGAGATCTCTGCG 606
167 aPheCysLeuThrPheIleGlyHis..... 175
607 TTCTGCTGTGTTCATAGCCATATAACTAAATCTATGCGATAATAG 656
175 ..... 175
657 GGCATATACCAGATCTCTGAGGCTTGTCGATTTCTATGTTATGCCGCTG 706
176 .....AsnArgMetTrrPar 180
707 GTTAATGTGACGGTTATGACTACTCATGAAAAAGAAATGAAATGTGAG 756
180 GlyAsnArgSerPheTYRAlaAsnAsnHisCysIleGlyThrAspLeuA 197
|||||
757 AAAGAACGTTCTTCTATGCGAAACAAATCATTTGATGGAACAGACTGA 806
197 sNArgAsnPheAlaSerLysHisTrrPcysGluGluGluValAlaSerSer 213
807 ATAGAACTTGTCTCCAAACACTGTGTGAGAGAGGTCATCCAGTTC 856
214 SerCysSerGluThrTYRCysGlyLeuTYRProGluSerGluProGluVal 230
|||||
857 TCATGCTCGGAAACCTACTGTGACTTATCTAGTCAGACAGACAGAGT 906
230 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaT 247
907 GAAGCAGGTGGCTAGTTCTTGAGAAAGAAATATACCCAGATTAAACAT 956
247 YRILeSerMetHisSerTYRSerGlnHisIleValPheProTYRSerTYR 263
957 ACATCAGCATGATTCATTACTCCAGCATATAGTGTTCATTAATTCAT 1006
264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
1007 ACAGGAAGTAAAGCAAAAGACCATGAGGAAGTGTCTGATAGCAGATGA 1056
280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTYRThrHisG 297
1057 AGCAGTTCTGCTATTGAGAAACTAGTAAATAATACAGATATACATG 1106
297 LysIleGlySerGluThrLeuTYRLeuAlaProGlyGlyLysAspTrrP 313
1107 GCCATGGCTCAGAAACCTTATTACTAGCTCCTGAGGAGGGGACGATGG 1156
314 IleTYRAspLeuGlyIleLysTYRSer.Phe..... 323
1157 ATCTATGATTTGGGCATCAATATATTCGTTTCAATTAAGCTGAGATAC 1206
324 .....ThrSerAsnProProVal 329
1207 GGGCACAATACGATTTCTGTGTCGCGGAGCGTTACATCAAAACCCACTGTA 1256
330 GluLysLeuLeuProLeuSerLeuLys 338
1257 GAGAAAGCTTTTGGCCGTGTCTCTATAAA 1283
```

seq_name: /cgn2_6/prodata/1/lna/5A.COMB.seq:US-08-696-139-1
seq_documentation_block:
; Sequence 1, Application US/08696139
; Patent No. 5672496

APPLICANT: Fayerman, Jeffrey T.

APPLICANT: Fayerman, Jeffrey T.
 APPLICANT: Greenen, David P.
 APPLICANT: Hersberger, Charles L
 APPLICANT: Larson, Jeffrey L.
 APPLICANT: Sterner, Jane L.

TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana

COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

· OPERATING SYSTEM: PC-DOS/MS-DOS
· SOFTWARE: PatentIn Release #1.0, Version #1.29
· CURRENT APPLICATION DATA:

FILING DATE: 08/08/2020 13:00
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:

FILING DATE: 16-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.

REFERENCE/DOCKET NUMBER: X-8681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1215 base pairs

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:

NAME/REF: CDS
LOCATION: 1..1215
08-696-139-1

gnment_scores:	
Quality:	646.50
Ratio:	2.694
length:	361
Gaps:	6

segment_block: 09-980-881-3 x US-08-696-139-1 ..

align seg 1/1 to: US-08-696-139-1 from: 1 to: 1215
1 pheclnSerglycInValleuAlaAlaLeuProArgThrSerArg

28 TTCGAAGGGGAGAAGGTCTTCCGTGTCATGTTGAAGATGAAATGACAT 77
17 LglnValLeuGlnAsnLeuThrThrTyrGluIleValLeuTyrGlnP 34

78 CAGCTTACTCCATGAGTTGGCCAGCACCAGCCAGATTGACTTCTGGAAC 127
34 rovalIThrAlaAspLeuIleValIysLysLysGlnValHisPhePheVal 50

128 CAGATTCGTCACCAAMTCMAACCTCACAGTACAGTTGACTTCCGTGTG 17
51 AsnAlaSerAspValaIspAsnValIlyAlaHisIleuAsnValSerGlyTl 67
...|||||...|||

178 AAAGCAGAGATATTTGGCTGTGGAAGACTTCTGAGCAGAATGAAC 227

228 ACAATATGAGGTACTCATTAACAACCTGAGATCTGTGCTCGAGGCTCAGT 277

278 TTGACAGCAGAGTC.....CGTACAACCTGGACACAGTTATGAGAAG 318

319 TACACAACTGGGAAACGATCGAGGCTTGAGCTAAGCAAGTCACCACGTGA 368

369 AATCCAGACCTCATCTCTCGCACGACCATCGGAATACATTTTGGAA 418

```

      ::|||:::|||||  ||||  |||
419 ACAATATATACCTCCTCAAGTT...GGCAAACCTGGACCATAAAGCCT 465

```

466 GCCATTTCATGAGACTGTGGTTTCCATGCGCAGAGATGGATTTCCCATGC 515

173 172

566 GTCACATGACAGATTCTCACAACAGCTAGACTTTATGTCTTGCCGTG 615
174G[V]H[is].....AsnArpMetTrnAr 180

616 CTCATATTGATGGCTACATCTACACCTGGACCAGAACCGAATGTGGAG 665
180 gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuA 197

666 AAGACCCGCTTACCAATGTGGAACTACCTGCATTGGCAGACACCCCA 715
197 snArgAsnPheAlaSerLyshHisTrpCysGluGluValAlaSerSerSer 213

716 ACAGAAATTTT...GATGCTGGGTGGTGACACACTGGAGCCTCTACAGAC 762
214 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVal 230

763 CCCTGCATGAGACCTTACTGTGGATCTGCTGCAGAGTCTGAAAAAGAGAC 812
230 LysAlaValAlaSerPheLeuArgArgAsnIleasnIleLysAlaI 247

247 yrIleSerMetHisSerTyrSerGlnHisIleValAlaPheProTyrSerTyr 263
813 CMAGGCCCTGGCTGNTTTATATACGCACACACCTCTCTCTCATCAAGCAT 862

264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
::: :::::

280 uAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisG 297

297 LYNISGLYSerGIuThrLeuTYrLeuAlaProGIuGIuAspSprTr 313

314 IleTyrAspLeuGlyIleLysTyrSerPheThr 324
|||||

[illegible][illegible]

93 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 109
|||||:::|||||::: ||| ::

67. GCACTGGTCCTCTTACGAGAAATCACACAGTGGGAAACATAGAGGC 116
109 rTTPleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 126
117 TTGGACGCAACAGTGGCCACGTGAGATTCAGCCCTCATCTCTCGCAGTG 166
126 tsIleGlySerSerPheGluLysTyProLeuTyValLeuLysValSer 142
167 TTATCGGAAACACATTTGAGGAGCGCGCTATTACCTCGAAGGCTT... 213
143 GlyLysGluInThrAlaLysAsnAlaIleTTPleAspCysGlyIleH 159
214 GGCAAAAGCTGGACAAAATTAAGCCTGCATTTTCATGAGCTGTGGTTCCA 263
159 sAlaArgGluTTPleSerProAlaPheCysLeuTTPheIle..... 173
264 TGGCAGAGATGGATTCTCTCGATTTCTGCCAGTGTGTTGTAAAGAGG 313
173 173
314 CTGTTCGTACCTATGACGTGAGATCCAGTACAGAGCTTCTGACAG 363
174 GlyHis..... 175
364 TTAGACTTTATGTCTGCTGCTGCTCAATATTGATGGCTACATCTACAC 413
176 AsnArgMetTTPArgLysAsnArgSerPheTyAlaAsnA 189
414 CTGGACCAAGCGCATTTTGGAGAAAGAGCTCGCTCCACCATCTGAT 463
189 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTTP 205
464 CTAGCTGCATTTGGCAGACGCCACAGAAATTTT... GATGCTGTTGG 510
206 CysGluGluInValAspSerSerSerCysSerGluThrTyCysGlyLe 222
511 TGTGATTTGGAGCTCTCGAAACCCCTGTGATGAACTTACTGTGGACC 560
222 uTyTProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 239
561 TGGCCGAGACTGTGAAAGAGAGACCAAGGCCCTGCGATTTTCACTCGCA 610
239 rGAsnIleAsnGlnIleLysAlaTyTleserMetHisSerTySerGln 255
611 ACAAACTCTCTCCATCAAGCATATCTGACAATCCACTGCTCCCAA 660
255 HisIleValPheProTySerTyTTPArgSerLysSerLysAspHisG 272
661 ATGATGATCTACCTTACTCATATGCTTACAAACTCGGTGAAGCAATGC 710
272 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 289
711 TGAGTTGAATGCCCTGCTAAGCTACTGTGAAGAAGCTT... GCCTCAC 757
289 eTyrAsnThrArgTyThrHisGlyHisGlySerGluThrLeuTTPleu 305
758 TGCACGCGACCAAGTACATATGCGCCGAGAGCTACACAACTATACCT 807
306 AlaProGlyLysGluAspTTPleTyTAspLeuLysIleLysValSer 322
808 GCTGCTGGGGGCTCTGACGACTGGCTTATGACCAAGAAATCAATATTC 857
322 rPheThr 324
858 CTTCAC 864

seq_name: /cgn2.6/plodata/1/ina/5B.COMB.seq:US-08-860-882A-64
seq_documentation_block:
; Sequence 64, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN

APPLICANT: EGGELTE, HENDRIKUS JOHANNES
APPLICANT: TARRAGONA-FIOL, ANTONIO
APPLICANT: RABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNAM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
APPLICANT: DAVIES, DAVID HUM
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 1053 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-64

alignment_scores:
Quality: 567.00 Length: 269
Ratio: 3.133 Gaps: 5
Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:
US-09-980-881-3 x US-08-860-882A-64 ..
Align seg 1/1 to: US-08-860-882A-64 from: 1 to: 1053

93 AlAserAlaSerTyTyGluGlnTyHisSerLeuAsnGlnIleTySer 109
117 TTGGACTGTACCTGTACGAGAAATACCAAGTGGGAAACGATAAGAGCC 116
109 rTTPleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 126
117 TTGGACTGTACCTGTACGAGAAATACCAAGTGGGAAACGATAAGAGCC 166
126 tsIleGlySerSerPheGluLysTyProLeuTyValLeuLysValSer 142
167 TTATCGGACCAACATTTGAGGAGCGCGTATTACCTCTCGTAAAGCTT... 213
143 GlyLysGluInThrAlaLysAsnAlaIleTTPleAspCysGlyIleH 159
214 GGCAAAAGCTGGACAAAATTAAGCCTGCATTTTCATGAGCTGTGGTTCCA 263
159 sAlaArgGluTTPleSerProAlaPheCysLeuTTPheIle..... 173
264 TGGCAGAGATGGATTCTCTCGATTTCTGCCAGTGTGTTGTAAAGAGG 313

```

1 ZIP: 20005
2 COMPUTER READABLE FORM:
3 MEDIUM TYPE: Floppy Disk
4 COMPUTER: IBM compatible
5 OPERATING SYSTEM: PC-DOS/MS-DOS
6 SOFTWARE: PatentIn Release #1.0, Version #1.25
7 CURRENT APPLICATION DATA:
8 APPLICATION NUMBER: US/08/860,882A
9 FILING DATE: JUNE 23, 1997
10 CLASSIFICATION: 424
11 ATTORNEY/AGENT INFORMATION:
12 NAME: DONALD J. BIRD
13 REGISTRATION NUMBER: 25,323
14 REFERENCE/DOCKET NUMBER: 9901/238653
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (202) 861-3027
17 TELEFAX: (202) 822-0944
18 TELEX: 6174627 CUSH
19 INFORMATION FOR SEQ ID NO: 74:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 1059 bases
22 TYPE: nucleic acid
23 STRANDEDNESS: single
24 TOPOLOGY: linear
25 US-08-860-882A-74
26
27 Alignment_scores:
28 Quality: 560.00 Length: 269
29 Ratio: 3.111 Gaps: 5
30 Percent Similarity: 66.914 Percent Identity: 40.892
31
32 alignment_block:
33 US-09-980-881-3 x US-08-860-882A-74 ..
34
35 Align seg 1/1 to: US-08-860-882A-74 from: 1 to: 1059
36
37 93 AlaseraIaSerTYTtYrGlugInThrHisSerLeuAsnGluIleYrse 109
38 GCAATCTGCTACTCTTACGAGAAAGTACAAACAGGGAAGATAGAGGC 116
39 rTripleGluThrHeIleThrGluThrGlnHisProAspMetLeuThrIleH 128
40 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
41 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
42 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
43 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
44 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
45 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
46 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
47 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
48 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
49 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
50 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
51 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
52 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
53 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
54 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
55 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
56 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
57 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
58 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
59 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
60 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
61 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
62 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
63 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
64 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
65 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
66 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
67 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
68 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
69 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
70 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
71 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
72 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
73 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
74 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
75 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
76 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
77 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
78 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
79 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
80 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
81 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
82 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
83 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
84 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
85 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
86 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
87 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
88 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
89 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
90 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
91 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
92 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
93 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
94 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
95 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
96 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
97 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
98 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
99 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
100 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
101 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
102 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
103 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
104 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
105 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
106 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
107 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
108 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
109 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
110 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
111 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
112 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
113 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
114 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
115 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
116 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
117 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
118 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
119 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
120 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
121 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
122 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
123 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
124 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
125 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
126 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
127 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
128 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
129 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
130 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
131 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
132 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
133 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
134 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
135 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
136 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
137 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
138 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
139 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
140 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
141 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
142 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
143 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
144 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
145 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
146 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
147 TTTGGCTCAACAAGTCGGCACTGGAATCCAGCCCTCATCTTCGCAAGT 166
148 TTTGGCTCAACAAGTCGGCACT
```

```

206 CysGluGluGlyAlaSerSerSerSerGluThrTyrcysgyle 222
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 TGGAATTGGAGCTCTCGAAACCCCTGATGAACCTTACGTGGACC 560
222 uTYrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 239
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 TGCCGAGAGCTGAAAGAGACCAAGGCCCTGGCTGATTTCGCGCA 610
239 rGAsnIleAsnGlnIleLysAlaIleTyrIleSerMetHisSerTyrSerGln 255
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
611 ACAACTCTCTTCATCAAGCATATCTGACAATCCAGCTGACTCCCAA 660
256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisgl 272
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 ATGATGATCTACCTTACTATATGCTTACCAACTCGGTGAGAACATGTC 710
272 uGluLeuSerLeuValAlaSerGluValAlaArgAlaIleGluLysThr 289
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
711 TGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAGAAGACT..GCCCTCAC 757
289 eRLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 305
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
758 TGCACGACCAAGTACATATGACATGCGCGGAGCTACACAACTATCTATCCT 807
306 AlaProGlyGlyLysAspTyrPheTyrAspLeuGlyLysTyr 322
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
808 GCGTCGGGGGCTTAAAGACTGGGCTTATGACCAAGAAATCAGATATTC 857
322 rPheThr 324
      |||||
858 CTTCACC 864

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-860-882A-77

```

seq_documentation_block:
; Sequence 77, Application US/08860882A
; Patent No. 5983281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUM
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944

```

```

; TELER: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1059 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-860-882A-77

alignment_scores:
      Quality: 559.00      Length: 269
      Ratio: 3.106      Gaps: 5
Percent Similarity: 66.914      Percent Identity: 40.892

alignment_block:
US-09-980-881-3 x US-08-860-882A-77
Align seg 1/1 to: US-08-860-882A-77 from: 1 to: 1059

93 AlaSerAlaSerTyrTyrGluGluGlnTyrHisSerLeuAsnGlnIleTyrSer 109
|||:||||| ||||| ||||| ||||| ||||| ||||| |||||
67 GCAACTGGTCACTCTTACGAAGATACACAAAGGGAAGCATAGAGGC 116
109 rTTPILeGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 126
|||:||||| ||||| ||||| ||||| ||||| ||||| |||||
117 TTGAGCTCAACAAGTCCGACCTGAGAAATCCAGCCCTCATCTCTGCACTG 166
126 IsIleGlySerSerPheGluLysTyrProLeuTyrValIleLysValSer 142
||||| ||||| ||||| ||||| ||||| ||||| |||||
167 TTATCGGAGACCACTTTGAGGAGCGCGCTATTATTCCTCTGAAAGTT... 213
143 GLYLysGluGlnThrAlaLysAsnAlaIleTTPILeAspCysGlyIleH 159
||||| ||||| ||||| ||||| ||||| ||||| |||||
214 GCGAAAGCTGCAAAATTAAGCCCTGCATTTTCATGAGCACTGTGTTCCA 263
159 sAlaArgGluTTPILeSerProAlaPheCysLeuTTPheIle..... 173
||||| ||||| ||||| ||||| ||||| ||||| |||||
264 TGCAGAGAGAGGATTTCTCCTGCATTCGCCAGCTGTTGTAAGAGAGC 313
173 ..... 173
314 CTGTTCCTACCTATGAGCTGAGATCCAAAGTACAGAGCTTTCGACAAAG 363
174 .....GlyHis..... 175
364 TTACACTTTTATGCTCGCTGCTGCTCAATATGATGCTACATCTACAC 413
176 .....AsnArgMetTyrPArgLysAsnArgSerPheTyrAlaAsnA 189
|||:||||| ||||| ||||| ||||| ||||| ||||| |||||
414 CTGAGCAAGAGCCGATTTTGGAAAGAGACCTGCCACCATCTGAT 463
189 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTTP 205
|||:||||| ||||| ||||| ||||| ||||| ||||| |||||
464 CTACCTGCATTGGCAGACAGCCCAAGAAATTTT...GATGCTGGTGG 510
206 CysGluGluGlyAlaSerSerSerSerSerGluThrTyrcysgyle 222
||||| ||||| ||||| ||||| ||||| ||||| |||||
511 TGGAATTGGAGCTCTCGAAACCCCTGATGAACCTTACGTGGACC 560
222 uTYrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 239
||||| ||||| ||||| ||||| ||||| ||||| |||||
239 rGAsnIleAsnGlnIleLysAlaIleTyrIleSerMetHisSerTyrSerGln 255
||||| ||||| ||||| ||||| ||||| ||||| |||||
611 ACAACTCTCTTCATCAAGCATATCTGACAATCCAGCTGACTCCCAA 660
256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisgl 272
||||| ||||| ||||| ||||| ||||| ||||| |||||
661 ATGATGATCTACCTTACTATATGCTTACCAACTCGGTGAGAACATGTC 710
272 uGluLeuSerLeuValAlaSerGluValAlaArgAlaIleGluLysThr 289

```


OM of: US-09-980-881-3 to: EST: * out_format : pfs

Date: Sep 18, 2002 5:41 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed-p2n.model -DEV=x1h
-O=/cgn2_1/USPRO.spool/US09980881/runat_16092002_140128_10623/app-query.fasta_1.1126
-DB=EST -QMT=fastap -SUFFIX=st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GCAPOP=4.500
-GCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCAPOP=6.000
-FCAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsom62 -TRANS=human40.cdl
-LIST=45 -DOCLALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=6 -MAXLEN=200000000 -USER=US09980881 @CGNL 1.3434
-NCPU=6 -ICPU=3 -IONCLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-3

Query length: 338

Database: EST: *

Database sequence: 13736207

Database length: -1841457050

Search time (sec): 3115.820000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_hhc:AK004045	+ 1494.00	2980.16	9.5e-157	1446	AK004045 Mus musculus 18 days
gb_est1:AL536069	+ 1244.50	2483.01	4.7e-129	889	AL536069 AL536069 IFLI.FL013.FBI
gb_est2:BI144644	+ 988.00	1966.63	2.7e-100	853	BI144644 602909932P1 NCI.CGAP.1
gb_est2:BG569281	+ 978.00	1946.49	3.6e-99	852	BG569281 60258873P1 NIH.MGC.76
gb_est2:BG616456	+ 926.00	1843.45	1.7e-93	738	BG616456 60261461BFI NIH.MGC.76
gb_est2:BI332659	+ 924.00	1832.58	5.4e-93	1022	BI332659 60298440BFI NIH.MGC.76
gb_est2:BG618629	+ 921.00	1832.51	8.0e-92	788	BG618629 60264606A1 NIH.MGC.76
gb_est2:BG565100	+ 912.00	1815.01	7.5e-92	747	BG565100 602583720P1 NIH.MGC.76
gb_est2:BG566051	+ 897.00	1784.78	6.6e-90	747	BG566051 60258252P1 NIH.MGC.76
gb_est1:AI048167	+ 894.50	1780.10	6.6e-90	724	AI048167 uc71b12.y1 Sugano mus
gb_est1:AV646979	+ 863.00	1716.46	2.3e-86	735	AV646979 AV646979 GLC Homo sapi
gb_est2:BG618239	+ 862.00	1714.60	2.9e-86	725	BG618239 602645567P1 NIH.MGC.76
gb_est1:AV651709	+ 831.50	1653.74	7.2e-83	688	AV651709 AV651709 GLC Homo sapi
gb_est2:BG567264	+ 828.50	1645.88	2.0e-82	806	BG567264 602589745P1 NIH.MGC.76
gb_est2:BG618813	+ 814.50	1618.50	6.6e-81	750	BG618813 60264618BFI NIH.MGC.76
gb_est1:AL536068	+ 812.50	1617.52	7.5e-81	575	AL536068 AL536068 IFLI.FL013.FBI
gb_est1:AI539872	+ 806.50	1603.83	4.4e-80	661	AI539872 uc13d11.y1 Sugano mus
gb_est1:AI182113	+ 806.00	1604.32	4.1e-80	580	AI182113 uc73f10.y1 Sugano mus
gb_est1:AA968307	+ 806.00	1604.15	4.2e-80	589	AA968307 uc71h07.y1 Sugano mus
gb_est2:BG568240	+ 785.00	1559.99	1.2e-77	691	BG568240 602587149P1 NIH.MGC.76
gb_est1:AW950700	+ 776.50	1544.43	8.9e-77	603	AW950700 EST362770 MAGE resseq
gb_est1:AI574117	+ 766.00	1525.18	1.0e-75	511	AI574117 uc68d09.y1 Soares fetal
gb_est1:AI558146	+ 759.00	1511.07	6.4e-75	511	AI558146 uc14a04.y1 Sugano mus
gb_est1:AV653037	+ 755.00	1501.50	2.2e-74	583	AV653037 AV653037 GLC Homo sapi
gb_est2:BG216536	+ 751.50	1490.20	9.3e-74	843	BG216536 RST36106 Althaysy RAGB
gb_est1:AV692032	+ 744.00	1478.96	3.9e-73	602	AV692032 AV692032 GLC Homo sapi
gb_est1:AV658330	+ 741.00	1472.94	8.5e-73	601	AV658330 AV658330 GLC Homo sapi
gb_est1:AV694058	+ 735.00	1460.43	4.2e-72	623	AV694058 AV694058 GLC Homo sapi
gb_est2:BG618950	+ 732.50	1457.10	6.5e-72	537	BG618950 za28c09.y1 Soares fetal
gb_est2:BG617911	+ 727.50	1443.20	3.9e-71	749	BG617911 60264656P1 NIH.MGC.76
gb_est1:AV698445	+ 725.00	1440.32	5.6e-71	621	AV698445 AV698445 GLC Homo sapi
gb_est1:AI332676	+ 715.50	1416.19	1.2e-69	958	AI332676 m124d06.y1 Soares mus
gb_est1:AI574064	+ 707.00	1404.26	5.7e-69	610	AI574064 uc75g11.y1 Sugano mus
gb_est1:AW413090	+ 698.50	1385.17	6.6e-68	723	AW413090 uc92c03.x1 Sugano mus
gb_est1:AI574057	+ 688.00	1368.36	5.5e-67	487	AI574057 uc75g04.x1 Sugano mus
gb_est1:AI325529	+ 685.50	1341.35	2.1e-66	772	AI325529 uc13d11.x1 Sugano mus
gb_est2:BF384322	+ 676.50	1331.22	1.8e-65	692	BF384322 602046670P1 NCI.CGAP.1
gb_hhc:AK003061	+ 667.00	1313.87	6.2e-64	1429	AK003061 Mus musculus adult ma
gb_est1:AI255397	+ 659.50	1306.21	1.6e-63	747	AI255397 uc14a04.y1 Sugano mus
gb_est2:W88434	+ 657.00	1306.87	1.5e-63	455	W88434 zn69f04.y1 Soares fetal

gb_est2:BI246210 + 656.50 1300.70 3.3e-63 713 BI246210 602958724P1 NCI.CGA
gb_est2:BG642020 + 649.00 1286.74 2.0e-62 645 BG642020 pgl11c.pK002.n16 Ch1
gb_est1:AI529266 + 642.00 1276.83 7.1e-62 448 AI529266 uc166h06.y1 Sugano m
gb_est1:AV658128 + 637.00 1263.11 4.1e-61 615 AV658128 AV658128 GLC Homo s
gb_est1:AB050299 + 631.00 1252.95 1.5e-60 520 AB050299 AB050299 lambda ZAP

seq_name: gb_hhc:AK004045

seq_documentation_block:

LOCUS AK004045 1446 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:1110032P04:catboxypeptidase B2 (plasma),
full insert sequence.

ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
clone_11b:RIKEN full-length enriched mouse cDNA library
clone:1110032P04.

ORGANISM

Mus musculus

REFERENCE

AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
METH. Enzymol. 303, 19-44 (1999)
JOURNAL 99279253
MEDLINE 10349636
PUBMED

REFERENCE

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL 20499374
MEDLINE 11042159
PUBMED

REFERENCE

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Taahito,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wachihi,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL 20530913
MEDLINE 11076861
PUBMED

REFERENCE

AUTHORS The RIKEN genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
MEDLINE

REFERENCE

AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carlandi,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Hoti,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasuawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koyu,S.,
Kurahara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Salto,H., Salto,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schiml,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tasaka,Y., Teijima,Y., Toya,T., Yamamura,T., Yamana,T.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M., and
Hayashizaki,Y.

REFERENCE

AUTHORS Direct Submission
TITLE Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

REFERENCE

JOURNAL


```

1155 ATCTATGATTTGGCATCAATATTTGTTTACATTTGAGCTCCGAGATAC 1204
324 .....ThisSerAnProProval 329
1205 AGCAGATACAGGATTTCTGCTGCTGAGAGATACATCAACACCACTTGAG 1254
330 GtUlysLeuLeuProLeuSerLeuLys 338
1255 CAGAGAGCTTTGGCCGCAATCTCTMAAA 1281

seq_name: gb_est1:AL536069

seq_documentation_block:
LOCUS AL536069 889 bp mRNA linear EST 13-FEB-2001
DEFINITION LTI_FL013_FBRn1 Homo sapiens cDNA clone CSODF022YE21 5
prime, mRNA sequence.
ACCESSION AL536069
VERSION AL536069.1 GI:12799562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 889)
AUTHORS L.H.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES
source
1..889
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODP022YE21"
/dev_stage="LTI_FL013_FBRn1"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
f.liang@lifestech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 255 a 186 c 192 g 255 t 1 others
ORIGIN

alignment_scores:
Quality: 1244.50 Length: 276
Ratio: 5.207 Gaps: 1
Percent Similarity: 86.594 Percent Identity: 86.232

alignment_block:
US-09-980-881-3 x AL536069 ..

Align seg 1/1 to: AL536069 from: 1 to: 889

1 PheGlnSerGlyValLeuAlaIleProArgThrSerArgIleVal 17
|||||
57 TTTTCAGAGTGGCCAGTCTTACGCTGCTTCTCTAGAACCTTAGGCAAGT 106
|||||
17 IGlValLeuGlnAsnLeuThrThrTyrgIuIleValLeuTrpGlnP 34
|||||
107 TCAAGTTCTACAGAAATCTTACTACACATATGAGATTGTTCTGTCGACG 156
|||||
34 rovalThrAlaSplLeuIleValLysLysGlnValHisPhePheVal 50

```

```

157 CGGTAAACGCTGACCTTATTTGTAGAGAAAAACAGTCATTTTGTGTA 206
|||||
51 AsnAlaSerAspValAlaSerValLysAlaHisLeuAsnValSerGlyI 67
|||||
207 AATGCATCTGATGTCCGACATGTGAAGCCCATTTTAATATGTAGCGGAAT 236
|||||
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
|||||
257 TCCATGAGATGCTTGTGCTGACAGCGTGGAAAGATCTATTCAACAGAGA 306
|||||
84 IeSerAsnAspThrValSerProArgAlaSerAlaSerTyrryGluGln 100
|||||
307 TTTTCAACGACACAGTCAGCCGCCGAGCCCTCCGATCTACTATGAACAG 356
|||||
101 TyrHisSerLeuAsnGlnIleTyrrSerTrpIleGluPheIleThrGlu 117
|||||
357 TATCACTCACTAAATGAATCTATTCTTGGATGATGAATTTTAACTGAGAG 406
|||||
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
|||||
407 GCATCTGATATGCTTACAAAATTCACATTTGATCTCTATTGAGAGAGT 456
|||||
134 yrrProLeuTyrrValLeuLysValSerGlyLysGluGlnThrAlaLys 150
|||||
457 ACCACCTCTATGTTTAAAGTTTCTGGAAAAAGAACAGCAGCCAAAAT 506
|||||
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerPro 167
|||||
507 GCCATATGATGATGACTGTGGAATTCATGCCAGAAATGATGATCTCTG 556
|||||
167 aPheCysLeuTrpPheIleGlyHis..... 175
|||||
557 TTTCTGCTGTGTGTTTCATAGCCATATTACTCAATTCTATGGGATATAG 606
|||||
175 ..... 175
607 GGCAGATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATGCGCGT 656
|||||
176 .....AsnArgMetTrpPar 180
|||||
657 GTTAAATGTGATGCTTATGACTCTACTATGAGAAAACGATGAATGTGAG 706
|||||
180 gLysAsnArgSerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeu 197
|||||
707 AAAGAACCGTTCTTTATGCAACATCATTCATGATGGAACAGACCTGA 756
|||||
197 snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 213
|||||
757 ATAGGAACCTTGTCTCCAAACAGTGTGAGAGAGGTGCATCCAGTTCC 806
|||||
214 SerCysSerGluThrTyrrCysGlyLeuTyrrProGluSerGluProGlu 230
|||||
807 TCATGCTCGGAACCTTACTGTGAGCTTTATCTGTGAGCAGACAGAGT 856
|||||
230 IlyAlaValAlaSerPheLeuArgArg 239
|||||
857 GAAGCAGTGTGCTAGTTTCTTGAGAGAA 884
|||||

seq_name: gb_est2:BI144644

seq_documentation_block:
LOCUS BI144644 853 bp mRNA linear EST 05-JUL-2001
DEFINITION 602909332F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5051358 5',
RNA sequence.
ACCESSION BI144644
VERSION BI144644.1 GI:14604645
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 853)

```

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L1M1139 row: 0 column: 07
 High quality sequence stop: 849.

FEATURES
 source
 Location/Qualifiers
 1..852
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="501358"
 /clone_lib="NCL_CGAP_L19"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: Liver; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: This is a NCL_CGAP library."

BASE COUNT 243 a 199 c 194 g 217 t
 ORIGIN

alignment_scores:
 Quality: 988.00 Length: 267
 Ratio: 4.638 Gaps: 4
 Percent Similarity: 79.775 Percent Identity: 72.659

alignment_block:
 US-09-980-881-3 x B1144644 ..

Align seg 1/1 to: B1144644 from: 1 to: 853

```

1  PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVa 17
|||||
59  TTTGAGATGGCGCAGTTTATCTGCTCTCCAGAACCTCCAGCAGT 108
|||||
17  GlnValLeuGlnAsnLeuThrThrTyrGluIleValLeuThrGlnP 34
|||||
109  TCAACTACTTCAGAACTTACTACAACTATGAGGCTCTCTGCGCAGC 158
|||||
34  rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
|||||
159  CAGTCACAGCTGAATTCATCGAAGAAAGAAAGAGTCCACTTTTGTGTG 208
|||||
51  AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly 67
|||||
209  AATCGCTGTGATGTCAGACGTCAAGCGCATTAATGTGAGCAGAAAT 258
|||||
67  eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
|||||
259  TCCATTAACTCTGATGAAACAACGTGAGACCTAATTAACAGCAGA 308
|||||
84  leSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGln 100
|||||
309  CTTTCATGACACGTCAGCCCGCCGCTCCCTTCTACTATAGACAG 358
|||||
101  TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 117
|||||
359  TATCACTCGCTAAATGAATCTATTCCTGATAGAAAGTCACTAAGTGA 408
|||||
117  gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluYst 134
|||||
409  GCATCTGACATGCTCCAGAAATCTACATCGATCATTCATTCAGAAAT 458
|||||
134  yrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150

```

```

|||||
459  ACCCACTTATGTTTAAAGTCTCAGAAAGAAAGAAAGATCAAAAAT 508
|||||
151  AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAl 167
|||||
509  GCCATCTGGATGACTGTGGAATCCATGCCAGAGATGATTCACCTGC 558
|||||
167  aPheCysLeuTrpPheIleGlyHis..... 175
|||||
559  TTTCGTGTTTGCTTCATAGGCTACGTACACAAATCCATGGGAAGAAA 608
|||||
175  ..... 175
609  ATCTGTATACAGACTCTGAGCAGGTGATTTCTACATATGCCCCGTG 658
176  .....AsnArgMetTrpPar 180
659  ATGAACGTGGATGCTATGACTACACGTGGAAGAAATCGAATGTGGAG 708
180  gLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 197
|||||
709  GAAGAACCCT...CTGTCAACAAGAACACCGTGGCTTGGCCACAGCTGA 755
197  snArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSer 213
|||||
756  ACAGGAACCTGCTCCAAACACTGTGTGAGAAAGTGCGCTCCAGTTCC 805
214  SerCysSerGlu..PhyTyrCys..GlyLeuTyrProGlnSerGluPro 228
|||||
806  TCTGCTGTGAAAACCTACTGTGGGACTTATCTGTGAGTGTGACCG 852

```

seq_name: gb_esl2:BG569281

seq_documentation_block:
 LOCUS BG569281 852 bp mRNA linear EST 10-APR-2001
 DEFINITION 602588573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5',
 mRNA sequence.
 ACCESSION BG569281
 VERSION BG569281.1 GI:13576934
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
 1 (bases 1 to 852)
 NIH-MGC <http://mgc.ncl.nih.gov/>.
 Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 CDNA Library Preparation: CLONETECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L1M1580 row: b column: 19
 High quality sequence stop: 751.

FEATURES
 source

Location/Qualifiers
 1..852
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4722354"
 /clone_lib="NIH_MGC_76"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: Liver; Vector: pCMV-LIB (Clontech); Site: 1: NotI (99cgtcgtgccc); Site: 2: SfiI (ggcattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAAGAGCGGAGGCGCGGACACATG-3' (30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 251 a 177 c 187 g 237 t
ORIGIN

alignment_scores:
Quality: 978.00 Length: 260
Ratio: 4.866 Gaps: 5
Percent Similarity: 77.308 Percent Identity: 75.385

alignment_block:
US-09-980-881-3 x BG569281 ..

Align seg 1/1 to: BG569281 from: 1 to: 852

```

135 ProleutyrPalleuysValserglyLysgluThrAlaLysAsnAl 151
|||||
2 CCACCTCATGTTTAAAGGTTCTGGAAAGAACACACCAAAATGC 51
|||||
151 alleTPLeaspCysglylleHsAlaArgLutPrpLseProAlap 168
|||||
52 CATATGATGACTGTGGATCCATGCCAGAAATGATCTCTCTGCT 101
|||||
168 heCysleutPrheilleglyHs..... 175
|||||
102 TCTGCTTGTTGCATAGGCCATATAACTCAATTCATGGATTAAGG 151
|||||
175 ..... 175
152 CAATATACCAATCTCTGAGGCTTGAGATTCTAGTATGCCAGTGT 201
|||||
176 .....AsnArgMetTrp.Arg 180
|||||
202 TAATGTGATGATTAGTACTACTCATGGAAGAAATCGAATGGAGCA 251
|||||
181 LysAsnArgSerPheTyraAlaAsnAsnHiscysIleglyThrAspLeuAs 197
|||||
252 AAGAACCGTTCTTTCATGCGAACAAATCATTCGATCGGAACAGACCTGAA 301
|||||
197 nAGAsnPhAlaSerlyHsIstrPCysgluLuglAlaSerSerSers 214
|||||
302 TAGAACATTGCTTCCAAACACAGTGTGTAGAGAGGTCATCCAGTCTCT 351
|||||
214 eCYSSerGluThrTyrcYsglyLeuTyrrProgluSerGluProgluVal 230
|||||
352 CATGCTCGGAACCTACTGTGACTTATCTGATGAGAACAGAACTG 401
|||||
231 LysAlaValAlaSerPheLeuArgAsnIleAsnIlelleLysAlaTy 247
|||||
402 AAGGAGTGCTAGTTCTTGAGAGAAATATCAACAGATTAAGCATTA 451
|||||
247 rIleSerMetHsSerTyrrSerGlnHsIleValPheProTySerTyrr 264
|||||
452 CATCAGCATCATTCATCTCCAGCATATAGTTCATATATTCCTATA 501
|||||
264 hrArgSerLySerLyAspHsIgluLusSerLeuValAlaSerGlu 280
|||||
502 CAGAGAGTAAAGCAAGAACATGAGAACTGTCTCTAGTAAAGCAAGTGA 551
|||||
281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrrHsIgl 297
|||||
552 GCAGTTCGTCCTATTCAGAAAATTAGTAAATAACAGTATACACATGC 601
|||||
297 yHsIglSerGluThrLeuTyrrLeuAlaProglyLysAspAspPrp 314
|||||
602 GCATGGGTGCAAACTTATACCTAGCTCTGGAGGTGGGAGCATTTGA 651
|||||
314 leTyrrAspLeuGlylleLysTyrrSerPheThr..... 324
|||||
652 TCATATGATTGGGACATACATATATGTTTACAAATGAACCTTCGAGATA 701
|||||

```

```

325 .....SerAsn..ProProva 329
702 CGGGACATTAACGAGATTCTTCGCCGAGCGGTACATCAAAACCACTGT 751
329 lGluLysLeuLeuProLeuSerLeu 337
|||||
752 AGAGAGCTTTGGCGGCTGTCTTA 776

```

seq_name: gb_est2:BG616456

seq documentation block: 738 bp mRNA linear EST 18-APR-2001
LOCUS BG616456
DEFINITION 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
mRNA sequence.
ACCESSION BG616456
VERSION BG616456.1 GI:13667827
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS NIH-MGC
TITLE 1 (bases 1 to 738)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bhs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINTL at:
http://image.llnl.gov
Plate: LLCM1595 row: c column: 05
High quality sequence stop: 666.

FEATURES
source
Location/Qualifiers
1..738
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4733500"
/clone_1lb="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); site_1:
3' adaptor were used in cloning as follows: 5' and
5' (ggcgctcgcc); site_2: 5' (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTAAGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 209 a 161 c 153 g 215 t
ORIGIN

alignment_scores:
Quality: 926.50 Length: 283
Ratio: 4.391 Gaps: 4
Percent Similarity: 74.558 Percent Identity: 67.845

alignment_block:
US-09-980-881-3 x BG616456 ..

Align seg 1/1 to: BG616456 from: 1 to: 738

```

1 PhgGlnSerGlyGlnValAlaLeuAlaLeuProArgThrSerArgGlnVal 17
|||||
40 TTTCAGAGTGGCCAAAGTTCAGTGTCTCTCTAGAACCTTGAAGCAAGT 89
|||||
17 lGlnValLeuGlnAsnLeuThrThrThrTyrrGlnIleValLeuTrpGlnP 34
|||||

```

```

90 TCAAGTTCTACAGATCTTACTACAAATATGAGATTGTTCTGCGACG 139
34 rovalThralaspLeuIleValLysLysLysGlnValHisPhePheVal 50
140 CGGTAAACACTGACCTTATTTGTAAGAAAAACAACAGTCCATTTTGTGA 189
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerLysI 67
190 AATGCATCTGATGTCGACAAATGGAAGCCCATTTAAATGTGAGCGCAAT 239
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
240 TCCATCAGTCTGCTTCTGCGACAGCTGGAAGATCTTATTCAACACACAG 289
84 leSerAsnAspThrValSerProAlaGalaSerAlaSerTyrTyrGluGln 100
290 TTTCCAAACACACAGTACAGCCCCCGAGCCCTCCGCAATGTAATGAGAG 339
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIlePheGlu 117
340 TATCACTCTCAATAATGAAATCTATTCTTGATGAAATTTATTAACAGAG 389
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
390 GCATCCTGATGCTTACAAAATCCACATGTGATCTCATTTGAGAGACT 439
134 yrrProLeuTyrValLeuLysValSerGlyLysGlnIleThrAlaLysAsn 150
440 ACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACACACACCAAAAT 489
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerPro 167
490 GCCATATGATGACTGCTGGAATCCATGCCAGAGATGATCTCTCTCC 539
167 aPheCysLeuTrpPheIleGlyHis..AsnArgMetTrpArgLysAsn 183
540 TTTCTGCTGTGCTCATAGGCCATATACATCAATCTATGGGATATATAG 589
183 gSerHeuTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 200
590 GCAATATATACCAATCTCTGAGGCTGTGTGA..... 620
200 heAlaSerLysHisTrpCysGluGlnGluAlaSerSerSerCysSer 216
621 .....TTTCTATGTTATGCCGGGTGTTCA..... 644
217 GluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 233
645 .....TGTGAGACTG..... 653
233 lAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSer 250
653 ..... 653
250 eHisLeuSerTyrSerGlnHisLeuAlaPheProTyrSerTyrThrArg 266
654 .....GTTATGAGACTACTCATGCGCAAGAAAT 680
267 LysSerLysAspHisGluGlnLeuSerLeuValAlaSerGluAlaVal 282
681 CGCACTGTGTGACAAAGAACCCGCTTACTAATGCAAGAAACAT 728

seq_name: gb_est2:BI332659
seq_documentation_block:
LOCUS      BI332659              1022 bp      mRNA      linear      EST 30-JUL-2001
DEFINITION 602984408f1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5137369 5',
ACCESSION  BI332659
VERSION     BI332659.1  GI:15017316
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus

```

```

REFERENCE
AUTHORS    Mammalia; Euthelia; Rodentia; Sciurognathi; Muridae; Mus.
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL    National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT    Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LRAM1337 row: 0 column: 02
High quality sequence stop: 841.

FEATURES
Source
Location/Qualifiers
1..1022
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5137369"
/clone_1lb="NCI_CGAP_L19"
/lab_host="DH10B (TI phage-resistant)"
/site_2="Organ: Liver; Vector: PCMV-SpOrf6; Site_1: NotI;
Average insert size 1.9 kb. Constructed by life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT      303 a      233 c      225 g      261 t
ORIGIN

alignment_scores:
Quality:      924.00      Length:      316
Ratio:        3.949      Gaps:        6
Percent Similarity: 74.051      Percent Identity: 61.709

alignment_block:
US-09-980-881-3 x BI332659 ..

Align seg 1/1 to: BI332659 from: 1 to: 1022

1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVal 17
|||||
60 TTTTCAGAGTGGCCAGCTTTATCTGCTCTCCAAAGACCTCAGGCAAGT 109
17 lGlnValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnP 34
|||||
110 TCACTACTTCAGAAATCTTACTACAAAGATGAGGTGCTCTGCGACG 159
34 rovalThralaspLeuIleValLysLysLysGlnValHisPhePheVal 50
|||||
160 CAGTGACAGCTGAATTCATCGAAGAAAGAAAGAACGTCACATTTTGTG 209
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerLysI 67
210 AATGCGTGTGATGTCGACAGTCTCAAGCGCATTTAAATGTGAGAGCAAT 259
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
|||||
260 TCCATTTAAGCTTGTGATGAAACAGAGTGAAGACCTTAATTGAACAGCA 309
84 leSerAsnAspThrValSerProAlaGalaSerAlaSerTyrTyrGluGln 100
|||||
310 CTTTCATATGACAGCTGAGCCCGCCGCTCCGTTCTATATAGAGAGAG 359
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 117
360 TATCACTCTCAATAATGAAATCTATCTCGATAGAGTCAATCACTAGACA 409
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLys 134
|||||

```

```

410 GCATCCTGCATGCTCCAGAAAATCTACATCGCATCTCTTGACAGACT 459
134 yPProleuTyValleuTyValSerGlyLysGluInThrAlaTyAsn 150
460 ACCACCTTATGTTTAAAGCTCTCAGGAAGAACAAAGAACAAAT 509
151 AlaIleTrrPleAspCysGlyIleHisAlaArgLutPrlSerProAl 167
510 GCCATCTGGATGCACTGTGAATCCATGCCAGGAATGATTCACCTGC 559
167 apheCysLeuTrrPheIleGlyHis 175
560 TTTCCTGTTC.TGGTTCAATAGGCTACGTGACACATTCATGCCAGAA 608
175 ..... 175
609 ATCTGATACCAGACTTCTGAGGACGTGGAATTTCTACATCATGCCCT 658
176 ..... AsnArgMetTrr 179
659 GATGGAACGTGATGGCTATGACTACACGTGACAAAAGAAATCGAATGTT 708
179 pArgTyAsnArgSerPheTyrrAlaAsnAsnHisCysIleGlyThr...A 195
709 GCAGGAAGAACCGTCTGCTCACCAGAACACCGCTGCTGTGGCAACATG 758
195 spleuAsnArgAsnPheAlaSerLysHisTrrPcysGluGluGlyAlaSer 211
759 ACCTGAACAGAACTTCGCTCCAAACACGTGTGTAAGTAAGTTGGCTC 808
212 SerSerSerCysSerGluInThrTyrrCysGlyLeuTyrr.ProGluSer.Glu 227
809 AAGTTTCCTCCTGCTCGAAACTGCTGACTTATTCCTCCAGACTGTAGAG 858
228 ProGluValLysAlaValAlaSerPheLeuArgAsnIleAsnGlnI 244
859 CCAGAACTGAAAGCCAGGGGTGACTTTTGAAGAAGATATCTGACCCCT 908
244 eTyrrAlaTyrrIleSerMetHisSerTyrrSerGlnHisIleValPhePro. 260
909 TAAAGTT.....ACTTCAGATGGCTCATATATCTCCAC 940
261 .....TyrrSerTyrrThrArgSerLysSerLysAspHisGlu 272
941 AATATGTTCTCTATTTTCATATGAAAGACGAGCCAGAA 984

seq_name: gb_ests2:BG618629

seq_documentation_block:
LOCUS BG618629 788 bp mRNA linear EST 18-APR-2001
DEFINITION 602646064F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
mRNA sequence.
ACCESSION BG618629
VERSION BG618629.1 GI:13670000
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 788)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgrabbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: ULCM1630 row: 0 column: 15

```

```

FEATURES             High quality sequence stop: 760.
source               Location/Qualifiers
                    1..788
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4767974"
                    /clone_11b="NIH_MGC_76"
                    /lab_host="DH10B (T1 phage-resistant)"
                    /note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:
                    SfiI (ggcgccctggcc); Site_2: SfiI (ggccatattggcc); 5' and
                    3' adaptors were used in cloning as follows: 5' adaptor
                    sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                    5'-ATTCTAGAGCCGAGCGGCGCAGCATG-dT(30)BN-3' (where B = A,
                    C, or G and N = A, C, G, or T). Average insert size 1.85
                    kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
                    by PCR. This library was enriched for full-length clones
                    and was constructed by Clontech Laboratories (Palo Alto,
                    CA). Note: this is a NIH-MGC Library."

BASE COUNT          222 a 157 c 169 g 240 t
ORIGIN

alignment_scores:
    Quality: 921.00      Length: 219
    Ratio: 5.088        Gaps: 1
    Percent Similarity: 82.648    Percent Identity: 81.735

Percent Similarity: 82.648    Percent Identity: 81.735

alignment_block:
US-09-980-881-3 x BG618629 ..

Align seg 1/1 to: BG618629 from: 1 to: 788

1  PheGlnSerGlyGlnValleuAlaAlaLeuProArgThrSerArgGlnVal 17
132 TTTGAGAGTGGCCAAAGTTTACGTAGCTGCTCTCTGAAACCTCTAGCAGT 181
17 1GlnValleuGlnAsnLeuThrThrThrTyrrGluIleValleuTrrPrl 34
182 TCAAGTTCTACAGAAATCTTACTACACATATGAAATGTTCTCTGCGAGC 231
34 roValThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
232 CGGTACACGCTGACCTTATTTGTGAAGAAAAACAAGTCCATTTTGTGTA 281
51 AsnAlaSerAspValAlaAspAsnValLysAlaHisLeuAsnValSerGly 67
282 AATGCACTGTGATGTCGACAAATGTGAAGCCCATTTAATGAGCGGAGAT 331
67 eProCysSerValleuLeuAlaAspValGluAspLeuIleGlnGlnI 84
332 TCCATGCAAGTGTCTTGGCTGGCAGATGTGAAGATCTTATTCACACAGAG 381
84 leSerAsnAspThrValSerProArgAlaSerAlaSerTyrrTyrrGluGln 100
382 TTTCACACGACACAGTCAGCCGCCGAGCCTCCGATCTGATATGAACAG 431
101 TyrHisSerLeuAsnGluIleTyrrSerTrrPrlleuPheIleThrGluAr 117
432 TATACATCAGTAATGAATCTATTTCTTGATTAATTAATTAATGAGAG 481
117 gHisProAspMetLeuThrLysLysIleHisIleGlySerSerPheGluLysT 134
482 GCATCTCTGATATGCTTACAAAATCCACATGATGATCCCATTTAGAGAGT 531
134 yPProleuTyrrValleuTyrrValSerGlyLysGluInThrAlaTyAsn 150
532 ACCCACTCTATGTTTAAAGGTTCTGGAAGAAAGAACAGCCCAAAAT 581
151 AlaIleTrrPleAspCysGlyIleHisAlaArgLutPrlSerProAl 167
582 GCCATATGATGATGACTGTGGAATCCATGCCAGGAATGATGATCTCTGCG 631
167 apheCysLeuTrrPheIleGlyHis..... 175

```

```

|||||
632 TTTCTCTGTTGGTTCATAGGCCATATMACTCATTTCTATGGATATAG 681
175 .....
682 GGCAATATTACCAATCTCGAGGCTTGATTTCTATGTTATGCAGT 731
176 .....
732 GGTAACTGTTGATGGTATGACTACTCATGGAAGAAAGCATCGATTG 781
179 PARGlys 181
|||||
782 GAGAAAG 788

seq_name: gb_est2:BG565100

seq_documentation_block:
LOCUS BG565100 747 bp mRNA linear EST 10-APR-2001
DEFINITION 602583720F1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4711348 5',
mRNA sequence.
ACCESSION BG565100
VERSION BG565100.1 GI:13572753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps+email.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CM1551 row: h column: 05
High quality sequence stop: 714.
Location/Qualifiers
1. 747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4711348"
/clone_1id="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
5'fl (ggcgccgcgcgc); Site_2: 5'fl (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCAGCGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 220 a 154 c 166 g 207 t
ORIGIN

```

```

alignment_scores:
Quality: 912.00 Length: 251
Ratio: 4.449 Gaps: 6
Percent Similarity: 81.673 Percent Identity: 78.088

```

```

alignment_block:
US-09-980-881-3 x BG565100 ..

```

```

Align seg 1/1 to: BG565100 from: 1 to: 747

```

```

30 valLeuTrpGlnProValThrAlaAspLeuIleValLysLysGlnIva 46
|||||
1 GTTCTGCGACGCGGTAAACAGTGCCTTAT. GTCAAAAAAACCAAGT 49
46 IHisPheIleValAsnAlaSerAspValAspAsnValLysAlaHisLeu 63
|||||
50 CCATTTTGTGTAGATGATCATGTGATGTCGACAAATGGAAGCCCATTTA 99
63 snValSerGlyIleProCysSerValLeuLeuAlaAspValGlnAspLeu 79
|||||
100 ATGTGAGCGGAATTCATCTCAGTCTTCTGCGCAGATGGAAGATCTT 149
80 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAla 96
|||||
150 ATTCAACACAGCATTTCCAAACGACACAGTCAAGCCCGGCGCTCCGCAAC 199
96 TrpTyrTrpGlnGlnIleTyrHisSerLeuAsnGlnIleTyrSerTrpIleGln 113
|||||
200 GTACTATGAAACAGTATCACTCACTAAATGAAATCTATTCTTGATAGAT 249
113 heIleThrGlnArgHisProAspMetLeuThrLysIleHisIleGlnSer 129
|||||
250 TTATTAAGTGAAGGATCTCTGATATGCTTACAAAATCCACATTTGATCC 299
130 SerPheGlnLysTyrProLeuTyrValLeuLysValSerGlyLysGlnI 146
|||||
300 TCATTTGAAAGTACCACTCATCTATGTTTAAAGTTTCGAAAAAGACA 349
146 nThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgIle 163
|||||
350 ACCAGCCAAAATATGCCATATGATGATTGACTGTGAAATCATATGCCAGAT 399
163 rPileSerProAlaPheCysLeuTrpPheIleGlyHis..... 175
|||||
400 GGACCTCTCCCTCTTCTGCTTGTGTTTCATAGGCCATATCACTCAATTC 449
175 ..... 175
450 TATGGATATATAGGCAATATACCAATCTCGAGGCTTGATGATTTCTA 499
176 .....
500 TGTATTGCCAGTGTAAATGTGATGTTATGACTCATCTAGGAAGCAAG 549
181 .....
550 AATCGGAATGTGAGACAGAAACCGTTCTTCTTATGCGAACATCATTTGC 599
192 IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGlnI 208
|||||
600 ATCGGAACAGACCTGAATAGGAACCTT. GCTTCCAAACACTGGTGTGAGCA 648
208 u.GlyAlaSerSerSerSerCysSerLeuThrTyrCysGlyLeuTyr 223
|||||
649 AGGTGATCCAGATTCCTCATTTGCTCGGAAAAACCTACTTGTGACTTTAT 698
224 ProGlnSerGlnProGlnValLysAlaValAlaSerPheLeuArg 238
|||||
699 CCTGAGTCAGAACCGAGAGTG. GAAGCAATGGCTACTTTCTTGAAAG 742

seq_name: gb_est2:BG566051

seq_documentation_block:
LOCUS BG566051 747 bp mRNA linear EST 10-APR-2001
DEFINITION 602582652F1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4710564 5',
mRNA sequence.
ACCESSION BG566051
VERSION BG566051.1 GI:13573704
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTTAAAGCTGGC and 3' end primer CGACCTGACCTGACGACA."

BASE COUNT 215 a 169 c 160 g 178 t 2 others
ORIGIN

alignment_scores:
Quality: 894.50 Length: 241
Ratio: 4.659 Gaps: 1
Percent similarity: 79.668 Percent identity: 70.954

alignment_block:
US-09-980-881-3 x AT048167 ..

Align seg 1/1 to: AT048167 from: 1 to: 724

```

1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVa 17
|||||
2 TTTCAGAGTGGCCAGCTTTTACTCTCTCCAGACACTCCAGCCACT 51
|||||
17 GlnValLeuGlnAsnLeuThrThrTyGluIleValLeuTrpGlnP 34
|||||
52 TCACCTACTTCAGAACTTACTACACAGTATGAGCTCTCTCGGACG 101
|||||
34 rovaltAlaAspLeuIleValLysLysGlnValHispheheVal 50
|||||
102 CAGTGACAGCTGAATTCATCCAGAGAAAGAGAGTCCACTTTTGTG 151
|||||
51 AsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGly 67
|||||
152 AATGGCTGTGATGTGACAGCTCAAGCGCATTTAAATGTGACAGAA 201
|||||
67 eProcysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 84
|||||
202 TCCATTATACGTTTGATGACACAGTGGAGGACCTTAATTGACGAGA 251
|||||
84 LeSerAsnAspThrValSerProArgAlaSerAlaSerTyTrpGln 100
|||||
252 CTTCACATGACACAGCTCAGCCCGCCCTCCCTTCACTATGACGAG 301
|||||
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 117
|||||
302 TATCACTCGCTAAATGAATCATCTCGATAGAGATCACTAATGACA 351
|||||
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGly 134
|||||
352 GCATCTGACATGCTCCAGAAATCTTACATCGATCATCTCGAAGT 401
|||||
134 YrProLeuTyValLeuLysValSerGlyLysGluGlnThrAlaLys 150
|||||
402 ACCCACTTATGTTTAAAGCTCTCAGACAGAGAAACAAGATCAAA 451
|||||
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerPro 167
|||||
452 GCCATCTGATGAGCTGTGGAATCATGTCAGAGAAATGATTTTCA 501
|||||
167 aPhcCysLeuTrpPheIleGlyHis..... 175
|||||
502 TCTCTGTTGTGGTTCAATGAGCTACGTACACAAATTCATGAGAG 551
|||||
175 ..... 175
552 ATCTGTATACAGACTTCTGAGGACAGTCTCTACATCATGACACTG 601
|||||
176 .....AsnArgMetTrpPar 180
|||||
602 ATGAACGTGATGGCTATGACTACACGTGAAAAAGAAATGCAATGTG 651
|||||

```

180 gLysAsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAspLeu 197
|||||
652 GAAGAGCCGCTGCTCACAAGAACACCGCTGCTGGGACAGACCTGA 701
|||||
197 snArgAsnPheAlaSerLysHis 204
|||||
702 ACAGGAACCTCGCTCCAAACAC 724
seq_name: gb_est1:AV646979

seq_documentation_block:

LOCUS AV646979 735 bp mRNA linear EST 15-JAN-2002
DEFINITION AV646979 GIC Homo sapiens cDNA clone GICATD06 3', mRNA sequence.
ACCESSION AV646979
VERSION AV646979.1 GI:9867993
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
1 (bases 1 to 735)
Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
Shen, K., Lu, G., Fu, G., Zhong, W., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z., and Han, Z.

Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106

COMMENT

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@cnhg.sh.cn

FEATURES

This clone is available at CHGC in Shanghai.
location/Qualifiers
1. 735

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GICATD06"
/clone_lib="GIC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 200 a 158 c 172 g 205 t
ORIGIN

alignment_scores:

Quality: 863.00 Length: 176
Ratio: 5.047 Gaps: 1
Percent similarity: 97.159 Percent identity: 96.591

alignment_block:
US-09-980-881-3 x AV646979 ..

Align seg 1/1 to: AV646979 from: 1 to: 735

```

1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnVa 17
|||||
58 TTTCAGAGTGGCCAGCTTTTACTCTCTCCAGACACTCCAGCCACT 107
|||||
17 GlnValLeuGlnAsnLeuThrThrTyGluIleValLeuTrpGlnP 34
|||||
108 TCAAGTCTCAGAACTTACTACACATATGAGATTTCTCTGGCAGC 157
|||||
34 rovaltAlaAspLeuIleValLysLysGlnValHispheheVal 50
|||||

```



```

158 CGGTAAACAGTGCACCTATTGTGAGAAAAAACAACTCCATTTTGTGA 207
51 AsnAlaSerAspValAspAsnValLysAlaHisLysAsnValSerGlyI 67
|||||
208 AATGCATCTGATGTCGACAAATGTGAAGCCCAATTTAAATGTGACGGAAT 257
67 eProcySerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnI 84
|||||
258 TCCATGCGACATGCTCTGCTGCGACAGCTGGAAGATCTTATTCAACACAG 307
84 leSerAsnAspThrValSerProArgAlaSerAlaSerIleTyrgluGln 100
|||||
308 TTTCACAGACACACAGACAGCCCCGAGCCTCCGCACTGGACATGAGACAG 357
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluar 117
|||||
358 TATCACTCTCAATGAATGAATCTATTCTTGATGATGAAATTTATACTAGAG 407
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysT 134
|||||
408 GCATCCTGATGATGCTTACAAAAATCCACATGATCCTCATTTGAGACAGT 457
134 TyrProLeuThrValLeuLysValSerGlyLysGluGlnThrAlaLysAs 150
|||||
458 ACCCACTCTATGTTTAAAGGGGTGTGAGAAAAGAACAAACAGCCAAAAA 507
150 nAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProa 167
|||||
508 TCCCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 557
167 lApheCysLeuTrpPheIleGlyHis 175
|||||
558 CTTTCTGCTGTGGGGGTCATAGGCCAT 583

seq_name: gb_est2:BG618239
LOCUS BG618239 725 bp mRNA linear EST 18-APR-2001
DEFINITION 602645967F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5',
mRNA sequence.
ACCESSION BG618239
VERSION BG618239.1 GI:13669610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 725)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
Plate: LICM1629 row: f column: 17
High quality sequence stop: 723.
Location/Qualifiers
1. 725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4767376"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1;
SfiI (ggcgccgcgcgc); Site: 2; SfiI (ggcgatctggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor sequence:

```

```

5'-ATCTAGAGCGCCGAGCGCGGCACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."
BASE COUNT 203 a 153 c 149 g 220 t
ORIGIN
alignment_scores:
Quality: 862.00 Length: 175
Ratio: 4.926 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.286
alignment_block:
US-09-980-881-3 x BG618239 ..
Align seg 1/1 to: BG618239 from: 1 to: 725
1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlna 17
|||||
131 TTTCAGAGTGGCCAGTTTACGCTGCTCTCCTAGAACCTTAGCGCAGT 180
17 lGlnValLeuGlnAsnLeuThrThrTyrgluIleValLeuTrpGlnP 34
|||||
181 TCAAGTTCTCTACAGAACTTCTACAACTATGAGAT.GTTCCTGCGCAGC 229
34 roValThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
|||||
230 CGGTAAACAGTGCACCTTAT.GTGAAGAAAAGAACAGTCTTTTGTGA 278
51 AsnAlaSerAspValAspAsnValLysAlaHisLysAsnValSerGlyI 67
|||||
279 AATGCATCTGATGTCGACAAATGTGAAGCCCAATTTAAATGTGACGGAAT 328
67 eProcySerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnI 84
|||||
329 TCCATGCGACATGCTCTGCTGCGACAGTGTGAAGATCTTATTCAACACAG 378
84 leSerAsnAspThrValSerProArgAlaSerAlaSerIleTyrgluGln 100
|||||
379 TTTCACAGACACAGTACGCCGCCGAGCCTCCGCACTGGACATGAGACAG 428
101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluar 117
|||||
429 TATCACTCTCAATGAATGAATCTATTCTTGATGAAATTTAATGAGACAG 478
117 gHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysT 134
|||||
479 GCATCCTGATGATGCTTACAAAAATCCACATGATCCTCATTTGAGACAGT 528
134 TyrProLeuThrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 150
|||||
529 ACCCACTCTATGTTTAAAGGGTTTCTGAAAAGAACAAACAGCCAAAAAT 578
151 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProa 167
|||||
579 GCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 628
167 lApheCysLeuTrpPheIleGlyHis 175
|||||
629 TTTCCTGCTGTGGGGGTCATAGGCCAT 653

seq_name: gb_est1:AV651709
LOCUS AV651709 688 bp mRNA linear EST 15-JAN-2002
DEFINITION AV651709 GIC Homo sapiens cDNA clone GLCSCF10 3', mRNA sequence.
ACCESSION AV651709
VERSION AV651709.1 GI:9872723
KEYWORDS EST.
SOURCE human.

```

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 688)
 AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, Y., Chen, Z., and Han, Z.
 Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

TITLE
 JOURNAL MEDLINE
 COMMENT 21625106
 Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@hgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1. 688
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="GLCCSF10"
 /clone_lib="GLC"
 /issue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 208 a 145 c 151 g 184 t

ORIGIN

alignment_scores:
 Quality: 831.50 Length: 181
 Ratio: 5.133 Gaps: 2
 Percent Similarity: 89.503 Percent Identity: 88.950

alignment_block:
 US-09-980-881-3 x AV651709 ..

Align seg 1/1 to: AV651709 from: 1 to: 688

176 AsnARgMetTrpArgLysAsnARgSerPheTyrlaAsnAsnHisCysII 192
 |||||
 46 AATCGAATGTGGAGAAAGAACCGTTCTTCATGCGAAACATCATCTGTCAT 95
 |||||
 192 eGlyThAspLeuAsnARgAsnPheAlaSerLysHisTrpCysGluGluG 209
 |||||
 96 CGGAACAGACCTGAAATAGAACTTCTCCAAACACTGGTGTAGAGAG 145
 |||||
 209 LysAlaSerSerSerCysSerGluThrTyrcysGlyLeuTyProGlu 225
 |||||
 146 GTGCATCAATCTCTCATGCTCGGAACCTACTGTGACCTTAACCGAG 195
 |||||
 226 SerGluProGluValLysAlaValAlaLaserPheLeuARgAsnIleAs 242
 |||||
 196 TCAGAACCAAGAGGAGGAGGAGTGGCTTCTTGAGAAAGAAATATCA 245
 |||||
 242 ngInIleLysAlaTyrlleSerMetHisSerTySerGlnHisIleValP 259
 |||||
 246 CCAGATTAAAGCATACATCAGATGCAATCTCATCTCCACACATTAAGT 295
 |||||
 259 heProTySerTyTrhArSerLysSerLysAspHisGluGluLeuSer 275
 |||||
 296 TTCATATCTCTATACACGAGTAAGCAAGACCATGAGGAACTGTCT 345
 |||||
 276 LeuValAlaSerGluValAlaValAlaIleGluLysThrSerLysAsnTh 292
 |||||
 346 CTAGTAGCCAGGAGGAGTTCGTCATATTGAGAAACCTAGTAAATATAC 395

292 rArgTyTrhHisGlyHisGlySerGluThrLeuTyTrhLeuAlaProGly 309
 |||||
 396 CAGGATACACATGGCCATGGCTCGAAGAACTTATACCTACTACTCTGAG 445
 |||||
 309 lYgLYAspAspTrpIleTyTrAspLeuGlyIleLysTySer.Phe.... 323
 |||||
 446 GTGGGAGCAGATTGGATTCATGATTTGGCGCATCAATATTCGTTACAA 495
 |||||
 324Thrs 325

496 GAACCTCGAGATACCGGCACATACCGATTCTTTGTGGCGGAGCGTTACAT 545
 |||||

325 eTAsnProProValGluLysLeuLeuProLeuSerIleLys 338
 |||||
 546 CAACCCACACCTGTGAGAGAGCTTTGGCGCTGTCTCTTAA 586
 |||||

seq_name: gb_est2:BG567264

seq_documentation_block:
 LOCUS BG567264 806 bp mRNA linear EST 10-APR-2001
 DEFINITION 602589745F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723690 5',
 mRNA sequence.
 ACCESSION BG567264
 VERSION BG567264.1 GI:13574917
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 806)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgrabs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCMI583 row: j column: 11
 High quality sequence stop: 772.

FEATURES
 source
 1. 806
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4723690"
 /clone_lib="NIH_MGC_76"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: liver; Vector: pNIR-LIB (Clontech); Site_1:
 SfiI (ggcgctcgcc); Site_2: SfiI (ggcctatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGCGAGCGCGCATATG-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC Library."

BASE COUNT 239 a 158 c 162 g 247 t

ORIGIN

alignment_scores:
 Quality: 828.50 Length: 179
 Ratio: 5.146 Gaps: 2
 Percent Similarity: 89.944 Percent Identity: 89.385

alignment_block:
 US-09-980-881-3 x BG567264 ..

Align seg 1/1 to: BG618264 from: 1 to: 806

```

178 MetTPARGLysAsnArgSerPheTYrAlaAsnAsnHisCysIleGly 194
    |||||||
2  ATGTGAGAAAGAACGCTCTTCTATGGACAAATCATTCATCGGAGAC 51
    |||||||
194 rAspleuAsnArgAsnPheAlaSerLysHisTrpCysGluGlyAla 211
    |||||||
52 AGACCTGAATAGAACCTTCTCCAAACACGCTGTGAGGAAGTCACAT 101
    |||||||
211 eSerSerSerCysSerGluThrTrpCysGlyLeuTyProGluSerGlu 227
    |||||||
102 CCAATTCCATGCTCGGAAACCTACTGTGGACTTATCTGTAGTCAGAA 151
    |||||||
228 ProGluValLysAlaValAlaSerPheLeuArgArgAsnIleasnGlu 244
    |||||||
152 CCAGAGTAGAAGCAGTGGCTAGTTCTTGAGAGAAATATCAACAGAT 201
    |||||||
244 eLysAlaTYrLleSerMetHisSerTYrSerGlnHisIleValPhePro 261
    |||||||
202 TAAAGCATATCATCAGCATCATCATCATCCACGACATATAGTGTTCAT 251
    |||||||
261 yPSeTYrThrArgSerLysSerLysAspHisGluGluSerLeuVal 277
    |||||||
252 ATTCCTATACAGAGATGAAAGCAAGACCATAGAGAACTGTCTAGTA 301
    |||||||
278 AlaSerGluAlaValAlaIleGlyLysTrpSerLysAsnThrArgTY 294
    |||||||
302 GCACAGTAAGCAGTGTGTGTATGAGAAACAGTAATAATACAGGTA 351
    |||||||
294 rThrHisGlyHisGlySerGluThrLeuTYrLeuAlaProGlyGlyAla 311
    |||||||
352 TACACATGGCCAGTGGCTCAGAAACCTATACCTAGCTCCGAGAGTGGGG 401
    |||||||
311 sPaSprPrlLeuTYrAspLeuGlyLleLysTYrSer.Phe..... 323
    |||||||
402 ACATTTGATCTATGATTTGGGCATCAATATTCGTTTCAATGAACCT 451
    |||||||
324 .....ThrSerAsn 327
    |||||||
452 CGAGATACGGGCACATACGATTCCTGCTCCGAGGAGTTACATCAAC 501
    |||||||
327 rProValGlyLysLeuLeuProLeuSerLeuLys 338
    |||||||
502 CACCTGTAGAGAGCTTTGCCCTGCTCTTAA 536

```

seq_name: gb_est2:BG618813

seq_documentation_block:

```

LOCUS      BG618813              750 bp      mRNA      linear      EST 18-APR-2001
DEFINITION 602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767885 5',
            mRNA sequence.
ACCESSION  BG618813
VERSION    BG618813.1 GI:13670184
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 750)
            NIH-MGC http://mgc.ncl.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
COMMENT    Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNU)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNU at:
            http://image.llnl.gov

```

Plate: L16M1630 row: k column: 22
High quality sequence stop: 622.
Location/Qualifiers

FEATURES
source

1. 750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4767885"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site: 1:
SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCCATTTGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCCAGGCGCGCAGCATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 219 a 146 c 171 g 213 t 1 others
ORIGIN

alignment_scores:
Quality: 814.50 Length: 188
Ratio: 4.735 Gaps: 5
Percent Similarity: 91.489 Percent Identity: 89.362

alignment_block:
US-09-980-881-3 x BG618813 ..

Align seg 1/1 to: BG618813 from: 1 to: 750

```

1 PheGlnSerGlyGlnValLeuAlaIleuProArgThrSerArgGlnVal 17
    |||||||
77 TTTCAGAGTGGCCAGATTCTAGCTGCTCTTCCTAGAACCTTAGGCAAGT 126
    |||||||
17 LglnValLeuGlnAsnLeuThrThrTYrGluIleValLeuTrpGlnP 34
    |||||||
127 TCAAGTTCTACAGATCTTACTACAAACATATGAGATGTTCTGTGCGACG 176
    |||||||
34 rovalThrAlaAspLeuIleValLysLysGlnValHisPhePheVal 50
    |||||||
177 CGGTAACAGTGCACCTTATGTGAGAGAAACAAAGTCCATTTTGTGTA 226
    |||||||
51 AsnAlaSerAspValAlaAspAsnValLysAlaHisLeuAsnValSerGly 67
    |||||||
227 AATGCATCTGATGCGACATGTGAAGCCCATTTAATGTGACGGGAAT 276
    |||||||
67 eProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnI 84
    |||||||
277 TCCATGCGAGTGTCTTGGCTGCGACAGCTGGAAGATCTTATTCACAGCAGA 326
    |||||||
84 leSerAsnAspThrValSerProArgAlaSerAlaSerTYrTYrGluGln 100
    |||||||
327 TTTCACAGACAGACAGTCAGACCCGAGCCCTCCGATGCTACTATGAAACAG 376
    |||||||
101 TYrHisSerLeuAsnGluIleTYrSerTrpIleGluPheIleThrGluAr 117
    |||||||
377 TATCAGCTACTAATGAATATCTATCTTGATAGAAATTTAATGTAGAGAG 426
    |||||||
117 gHisProAspMet.LeuThrLysIleHisIleGly.SerSerPheGlyLys 133
    |||||||
427 GCATCCGTGATATGCTTAACAAAATCCACATTTGCGATCCATCATCGAGAA 476
    |||||||
133 sTYrProLeu.TyrValLeuLysValSer.GlyLysGluGlnTrpAla 149
    |||||||
477 GTAACAGTGTATGTTTAAAGGTTCTCTGGGAAACAAACAGAGGCCAA 526
    |||||||
149 sAsnAlaIleTrpLleAspCysGlyIleHisAlaArgGluTrpLleSerP 166
    |||||||
527 AATATGCATATGATGTAGCTGTGGAATCCATCCAGAGAAATGATCTCTC 576

```

166 roAlaPheCysLeuTrp.....PheIleGlyHis 175
|||||
577 CTGCTTCTGCTGTGTGCATAGGCCATATAACTCAAAATTCTATGGGAT 626
176 AsnArg 177
|||||
627 AATAGG 632

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on: September 18, 2002, 16:47:50 : Search time 23.83 seconds  
          (without alignments)  
        399.707 Million cell updates/sec
```

Title: US-09-980-881-4
Page: 1320

Sequence: 1 ASASYEQLHSLNEIYSWIE.....IKYFTSNPVEKLLPLSLK 246

Scoring table: BLOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

```
Minimum DB seq length: 0
```

Maximum DB seq length: 20000000000

Post-processing:	Minimum Match	0%
------------------	---------------	----

Listing first 45 summaries

Database :

SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	601	44.9	415	1	CBP2_RAT	P19233 rattus norv
2	586	43.8	306	1	CBP2_BOVIN	P00732 bos taurus
3	578	43.2	416	1	CBP2_CANFA	P55261 canis famli
4	575	43.0	401	1	CBP2_PIG	P09955 sus scrofa
5	567	42.4	417	1	CBP2_HUMAN	P15086 homo sapien
6	561	41.9	417	1	CBP2_MOUSE	P15089 mus musculu
7	552	41.3	309	1	CBP2_RAT	P21661 rattus norv
8	552	41.3	417	1	CBP2_HUMAN	P15088 homo sapien
9	502.5	37.6	419	1	CBP2_BOVIN	P00730 bos taurus
10	488.5	36.5	419	1	CBP1_RAT	P00731 rattus norv
11	483	36.1	417	1	CBP2_RAT	P19222 rattus norv
12	465.5	34.8	419	1	CBP1_HUMAN	P15085 homo sapien
13	458	34.2	421	1	CBP2_HUMAN	P09142 homo sapien
14	456	34.1	417	1	CBP2_HUMAN	P48052 homo sapien
15	411.5	30.8	304	1	CBP2_STMVI	P427188 simulium v
16	381.5	28.5	433	1	CBP2_ANOGA	P002350 anopheles
17	375.5	28.1	303	1	CBP2_ASEFI	P04069 astacus flu
18	286.5	21.4	424	1	CBP2_THERV	P29068 thermocactin
19	286.5	21.4	420	1	YH72_YERST	P38836 saccharomyc
20	282.5	21.1	451	1	CBP5_STRGR	P18144 streptomyc
21	255	19.1	434	1	CBP5_STRCP	P39041 streptomyc
22	103.5	7.7	445	1	TR5H_CHICK	P70080 gallus gari
23	99	7.4	444	1	TR5H_HUMAN	P17752 homo sapien
24	97	7.2	297	1	Y103_HUMAN	P15062 homo sapien
25	96.5	7.2	376	1	YQCT_BACSU	P54497 bacillus su
26	95	7.1	444	1	TR5H_RABIT	P54490 oryctolagus
27	95	7.1	767	1	ORC1_SCHPO	P17750 schizosacch
28	91	6.8	202	1	YC01_ECOLI	P54789 escherichia
29	90.5	6.8	434	1	CBP8_BOVIN	P04836 bos taurus
30	90.5	6.8	476	1	CBP8_RAT	P15087 rattus norv
31	89.5	6.7	426	1	CBP8_MOUSE	P00493 mus musculu
32	89	6.7	461	1	TR5H_XENLA	P092142 xenopus lae
33	88.5	6.6	476	1	CBP8_HUMAN	P16870 homo sapien

	ID	CBPB_RAT	STANDARD;	PRT:	415 AA.	
RESULT_1						
CBPB_RAT						
p19223;						
DJ	01-NOV-1990	(Rel. 16,	Created)			
DT	01-NOV-1980	(Rel. 16,	Last sequence update)			
DE	16-OCT-2001	(Rel. 40,	Last annotation update)			
GN	Cardoxypeptidase B precursor	(EC 3.4.17.2).				
OS	Rattus norvegicus (Rat) .					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
NL	NCBI_TaxId=10116;					
RA	[1]					
RL	SEQUENCE FROM N.A.					
CC	MEDLINE=89034324; PubMed=3182872;					
RT	Clausner E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;					
FT	"Structural characterization of the rat carboxypeptidase A1 and B					
TT	genes. Comparative analysis of the rat carboxipeptidase gene					
J.F. Biol. Chem.	263:17837-17845(1988).					
-I-	CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =					
-I-	peptide + L-Hisnne(or L-arginine).					
ZINC CARBOXYPEPTIDASE FAMILY M14; ALSO KNOWN AS THE						
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@lsb.slb.ch).						
EMBL; M23959; AAAA0872.1; JOINED.						
EMBL; M23947; AAAA0872.1; JOINED.						
EMBL; M23950; AAAA0872.1; JOINED.						
EMBL; M23952; AAAA0872.1; JOINED.						
EMBL; M23953; AAAA0872.1; JOINED.						
PIR; A32129; A32129.						
HSSP; P09955; INSA.						
MEROPS; M14.003; --						
IProto; IPR003146; Propep_M14.						
IRefProt; IPRO00834; Zn_carbopept.						
Pfam; PF002244; Propep_M14; 1.						
Pfam; PF00245; Zn_carbopeptase; 1.						
PRINTS; PR00765; CRBOXYPEPTASE.						
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.						
KW PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.						
Hydrolase; Carboxypeptidase; Metalloproteinase; zinc; Zymogen; Signal.						
SIGNAL	1	13				
PROPSEP						
CHARIN	14	108				
CHARIN	109	415				
CHARIN	174	174				
METAL	177	177				

ALIGNMENTS

p22809 dtrosophila
 P17532 mus musculus
 P27306 escherichia
 P42787 dtrosophila
 O1505 homo sapien
 P27206 bacillus su
 P52906 caenorhabdi
 P09810 rattus norv
 P37892 lophius ame
 P318890 saccharomyc
 O42091 anguilla am
 P15169 homo sapien

FT METAL 302 302 ZINC (BY SIMILARITY).
 FT ACT_SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 171 184 BY SIMILARITY.
 FT DISULFID 243 266 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 SQ SEQUENCE 415 AA; 47515 MW; 8EA06CCADE30B6F2 CRC64;

Query Match 44.9%; Score 601; DB 1; Length 415;
 Best Local Similarity 44.6%; Pred. No. 6.8e-49;
 Matches 120; Conservative 42; Mismatches 67; Indels 40; Gaps 5;

QY 1 ASASYEQYSLNFIYSLIEFITERHPDMLTKIHGSEFEKYPYLKVSKEQTANNAI 60
 DB 109 ASGSHYTKYMETEAMIQOAVATDNDPLVQSVIGTTFEGRNMYLKI-CKTRPKPAI 167
 QY 61 WIDGHHAREMISPAFLMFI-----GH----- 83
 DB 168 FIDCGFHAREMISPAFCOMFVREAVRTYNOEIHMKQLDELDFYVLPPVNIIDGYVTWTK 227
 QY 84 NRMWRKNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVAS 143
 DB 228 DMRWRKTRSTWAGSSCGVPRNRNF-NAGMCEVGAASPCSEYTCGPAPESEKETKALAD 286
 QY 144 FLRNNIQIAVISMHSYQHIVFPYSTRSKSKDHELSLVAASEAVRAIEKTSKNTRYT 203
 DB 287 FIRNNLSITIAVILHYSYQMMLYPSYDYKLPENYEELNALVGAARKEL-ATLHGTRKT 345
 QY 204 HGHSEETLYLAPGGDDMIYDLGIRYSFT 232
 DB 346 YGPGATTIYPAAGSGDDMSYDGIKIRYSFT 374

RESULT 2
 ID CBPB BOVIN STANDARD; PRT; 306 AA.
 AC P00732;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B (EC 3.4.17.2).
 GN CPB.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RP SEQUENCE.
 RX MEDLINE=75217824; PubMed=1057162;
 RA Tilani K., Ericsson L.H., Walsh K.A., Neurath H.;
 RT Amino-acid sequence of bovine carboxypeptidase B.";
 RT Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
 RN [2]
 RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
 RX MEDLINE=74260705; PubMed=4833744;
 RA Schmidt J.U., Hirs C.H.W.;
 RT Primary structure of bovine carboxypeptidase B. Inferences from the
 RT locations of the half-cysteines and identification of the active site
 RT arginine.";
 RT J. Biol. Chem. 249:3756-3764(1974).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
 RX MEDLINE=76265065; PubMed=957425;
 RA Schmidt M.F., Herriott J.R.;
 RT Structure of carboxypeptidase B at 2.8-A resolution.";
 RT J. Mol. Biol. 103:175-190(1976).
 RN [4]
 RP ACTIVE SITE.
 RX MEDLINE=70007159; PubMed=5344132;
 RA Plummer T.H., Jr.;
 RT Isolation and sequence of peptides at the active center of bovine
 RT carboxypeptidase B.";

RL J. Biol. Chem. 244:5246-5253(1969).
 RN [5]
 RP ACTIVE SITE.
 RX MEDLINE=73061487; PubMed=4565668;
 RA Kimmel M.T., Plummer T.H., Jr.;
 RT Identification of a glutamic acid at the active center of bovine
 RT carboxypeptidase B.";
 RT J. Biol. Chem. 247:7864-7869(1972).
 CC -1 CATALYTIC ACTIVITY: peptidyl-L-lysine(or L-arginine) + H(2)O =
 CC peptide + L-lysine(or L-arginine).
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC PIR: A00912; CPEOB.
 DR PDB: 1CPB; 30-SEP-83.
 DR MEROPS: M14.003; -.
 DR InterPro: IPR00834; Zn_carboxypeptidase.
 DR Pfam: PF00246; Zn_carboxypeptidase.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
 FT DISULFID 63 76
 FT DISULFID 135 158
 FT METAL 149 163
 FT METAL 66 66 ZINC.
 FT METAL 69 69 ZINC.
 FT METAL 194 194 ZINC.
 FT ACT_SITE 246 246
 FT ACT_SITE 268 268
 SQ SEQUENCE 306 AA; 34612 MW; C329D265C44A172 CRC64;

Query Match 43.8%; Score 586; DB 1; Length 306;
 Best Local Similarity 43.9%; Pred. No. 1.2e-47;
 Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;

QY 6 YEQYHSLENIYSLIEFITERHPDMLTKIHGSEFEKYPYLKVSKEQTANNAIWIDCG 65
 DB 6 YEKNNMETEAMIEQVASENPDLISRAIGTFLGNTIYLLVY-GPGSNKPAVFDG 64
 QY 66 IHAREMISPAFLMFI-----GH-----NRMWR 88
 DB 65 FHAREMISPAFCOMFVREAVRTYNGREIHMTEFLDKLDFYVLPPVNIIDGYVTWTNMMR 124
 QY 89 KNSFYANNHCIGTDLNRNFASKHMCCEGASSSCSEYTCGLYPESEPEVKAVASFLRN 148
 DB 125 KTRSTRAGSSCTGDLNRNF-DAGWCISGASNNPCSEYTCGSAASESEKKAAYADFI RNH 183
 QY 149 INQIKAVISMHSYQHIVFPYSTRSKSKDHELSLVAASEAVRAIEKTSKNTRYTHGHS 208
 DB 184 LSSIKAVLTHYSYQMMLYPSYDYKLPKNVVELNLAKGAVKKL-ASLHGTTYSYGP 242
 QY 209 EETLYLAPGGDDMIYDLGIRYSFT 232
 DB 243 TTIYPASGSGDDMAYDGIKIRYSFT 266

RESULT 3
 ID CBPB CANFA STANDARD; PRT; 416 AA.
 AC P55261;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
 DE membrane associated protein) (ZAP47).
 GN CPB1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OC NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE-Pancreas;
 RA Fukuoka S.-I.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
 CC Peptide + L-Lysine(or L-arginine).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D78348; BAA1366.1; .
 CC HSSP: P09955; 1PBA.
 DR MEROPS: M14.003; .
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxept; 1.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 109 ACTIVATION PEPTIDE (BY SIMILARITY).
 FT CHAIN 110 416 CARBOXYPEPTIDASE B.
 FT METAL 175 175 ZINC (BY SIMILARITY).
 FT METAL 178 178 ZINC (BY SIMILARITY).
 FT METAL 303 303 ZINC (BY SIMILARITY).
 FT ACT_SITE 377 377 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 244 267 BY SIMILARITY.
 FT DISULFID 258 272 BY SIMILARITY.
 SQ SEQUENCE 416 AA; 47595 MW; DF1992CD52F8AB4 CRC64.

Query Match 43.2%; Score 578; DB 1; Length 416;
 Best Local Similarity 42.4%; Pred. No. 9.8e-47;
 Matches 114; Conservative 46; Mismatches 69; Indels 40; Gaps 5;

OY 1 ASASYEYDYSINETYSWETTERHPDMLTIHIGSSSEKPLVLYLAKSGEQAKNAI 60
 DB 110 ATGHSEYKXNWEETLEATQOVTSENPDISRSIGTTEGRTIYLAKY-GRAGQKPAI 168
 OY 61 WIDCGIHAREWISPAFCLEFI-----GH----- 83
 DB 169 FMDGCFHAEWISPAWQWFEVREKXRTYGOEIHMTLKDLPYVLPVGNIDGYVTTWK 228
 OY 84 NNMWRKNSFYANNHICIGTDLNRFASKHCEGASSSSCSEYCGLYEPESEPEVAVAS 143
 DB 229 NNMWRKSTSTQVTCNCVGTDPTRNF-DAGWCKIGASRNCCDEYTCPPAASESEKETAALN 287
 OY 144 FLRRNINOIKAVISMHSQHLVPEFYSYTRSKDHELSIVASEAVRAIEKTSKNTRYT 203
 DB 288 FLRSLNLSIKAVITLHSYSQMLPYSDYKLTENNNAELNALAKATVKEL-ATLHGTXYT 346
 OY 204 HGHSEETLYLAPGGDDWTYDGIKYSFT 232
 DB 347 YGPGATTYTPAAGSGDDWATVDOGIKYSFT 375

OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE OF 1-103.
 RX MEDLINE=91208150; PubMed=2018774;
 RA Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
 RA Aviles F.X.;
 RT "Analysis of the activation process of porcine procarboxypeptidase B
 RT and determination of the sequence of its activation segment.";
 RL Biochemistry 30:4082-4089(1991).
 RN [2]
 RP PRELIMINARY SEQUENCE OF 1-38.
 RX MEDLINE=85279427; PubMed=4026847;
 RA Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
 RT "Sequential homologs between procarboxypeptidases A and B from
 RT porcine pancreas.";
 RL Biochem. Biophys. Res. Commun. 130:97-103(1985).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE=91114690; PubMed=1989878;
 RA Coll M., Guasch A., Aviles F.X., Huber R.;
 RT "Three-dimensional structure of porcine procarboxypeptidase B: a
 RT structural basis of its inactivity.";
 RL EMBO J. 10:1-9(1991).
 RN [4]
 RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
 RX MEDLINE=9102767; PubMed=2223783;
 RA Vendrell J., Wider G., Aviles F.X., Wuehrich K.;
 RT "Sequence-specific 1H NMR assignments and determination of the
 RT secondary structure for the activation domain isolated from
 RT pancreatic procarboxypeptidase B.";
 RL Biochemistry 29:7515-7522(1990).
 RN [5]
 RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
 RX MEDLINE=91114693; PubMed=1989879;
 RA Vendrell J., Billeter M., Wider G., Aviles F.X., Wuehrich K.;
 RT "The NMR structure of the activation domain isolated from porcine
 RT procarboxypeptidase B.";
 RL EMBO J. 10:11-15(1991).
 RN [6]
 RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
 RX MEDLINE=93044373; PubMed=1422143;
 RA Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
 RA Huber R., Wuehrich K.;
 RT "Comparison of the NMR solution structure with the X-ray crystal
 RT structure of the activation domain from procarboxypeptidase B.";
 RL J. Biomol. NMR 2:1-10(1992).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
 CC Peptide + L-Lysine(or L-arginine).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -1- DATABASE: NAME-Worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/C/COB.html".
 CC PIR: B29181; B29181.
 DR PDB: 1PBA; 31-OCT-93.
 DR PDB: 1NSA; 24-DEC-97.
 DR MEROPS: M14.003; .
 DR InterPro: IPR003146; Propep_M14.
 DR InterPro: IPR000834; Zn_carboxept.
 DR Pfam: PF02244; Propep_M14; 1.
 DR Pfam: PF00246; Zn_carboxept; 1.
 DR PRINTS: PR00765; CRBOXYPTASEA.
 DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
 KW 3D-structure.
 FT PROPEP 1 95 ACTIVATION PEPTIDE.
 FT CHAIN 96 401 CARBOXYPEPTIDASE B.
 FT DISULFID 158 171
 FT DISULFID 230 253
 FT DISULFID 244 258

```

FT METAL 161 161 ZINC.
FT METAL 164 164 ZINC.
FT METAL 289 289 ZINC.
FT ACT_SITE 341 341
FT ACT_SITE 363 363 NUCLEOPHILE.
FT STRAND 14 17
FT HELIX 20 31
FT TURN 32 32
FT HELIX 43 45
FT STRAND 50 52
FT HELIX 61 69
FT TURN 70 72
FT STRAND 75 76
SQ SEQUENCE 401 AA; 45713 MW; 53129AF159A26348 CRC64;

```

```

Query Match 43.0%; Score 575; DB 1; Length 401;
Best Local Similarity 43.6%; Pred. No. 1.8e-46;
Matches 115; Conservative 45; Mismatches 64; Indels 40; Gaps 5;

```

```

OY 6 YEYHSLNEIYSWIEFTEHRPDM.LTKIHSSFEKYP.LYLVKSGKQOTAKNAIWDG 65
D 101 YEKYNNMETTIAWTEQVYKSKNDLISRSATIGTFDGNIVILKV-GKFGSKNPALFMDG 159
OY 66 IHAREWISPARCLWFI-----GH-----NRMR 88
D 160 FHAREWISQACQWVFAVRYTEYEAHMEFELNDLFYVLPVINDIGYITWTNRMR 219
OY 89 KNRSPYANNHCIGDNLNRFASKHMCCEGASSSCSETYCGLYPEPEVAVASFRLRN 148
D 220 KTRSTNNGSSCTGTDPNRFN-NAQMCYTGASVNPNCNEYGSAASEKETALADFTINN 278
OY 149 INOKATISMHSYQHIYFPEYSTRSKSKDHEELSLVASEAVRAIEKTSKTRTYHGS 208
D 279 LSSIKAVLTHYSQOMLTPYSDYKLPENDAEELSLAKAVKEL-ASLYGTSYSGPGS 337
OY 209 ETLVAPGGDDMIYDGIKISFT 232
D 338 TTITPAAGSGDDMAVYNGIKISFT 361

```

```

RESULT 5
CBP_HUMAN STANDARD; PRT; 417 AA.
AC p15086: 060834;
DT 01-APR-1990 (Rel. 14, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein)
(PASP).
GN CPB1 OR CPB OR PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC TISSUE=Pancreas; PubMed=1370825;
RX MEDLINE=92129345; PubMed=1370825;
RA Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA French C.K.;
RT "Isolation of a cDNA encoding a human serum marker for acute
pancreatitis. Identification of pancreas-specific protein as
pancreatic procarboxypeptidase B."
RL J. Biol. Chem. 267:2575-2581(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas; PubMed=9524066;
RX MEDLINE=98182241; PubMed=9524066;
RA Aloy P., Catusas L., Villegas V., Reverter D., Vendrell J.,
RA Aviles F.X.;
RT "Comparative analysis of the sequences and three-dimensional models
of human procarboxypeptidases A1, A2 and B."

```

```

RL Biol. Chem. 379:149-155(1998).
RN [3]
RP SEQUENCE OF 16-43.
RC TISSUE=Pancreas;
RX MEDLINE=89153096; PubMed=2920728;
RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT "Purification and properties of five different forms of human
procarboxypeptidases."
RL Eur. J. Biochem. 179:609-616(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2)O =
peptide + L-lysine(or L-arginine).
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81057; AAA66973.1; -
DR EMBL; AJ224866; CAA12163.1; -
DR PIR; S02812; S02812.
DR PIR; A42332; A42332.
DR HSSP; P09955; INSA.
DR MEROPS; M14.003; -.
DR MTM; 114852; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_Carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_Carboxypept. 1.
DR PRINTS; PR00765; CARBOXYPEPTASE.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR KMW; Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 110
FT CHAIN 111 417
FT METAL 176 176
FT METAL 179 179
FT METAL 304 304
FT ACT_SITE 378 378
FT DISULFID 173 186
FT DISULFID 245 268
FT DISULFID 259 273
FT CONFLICT 16 16
FT CONFLICT 17 17
FT CONFLICT 37 37
FT CONFLICT 208 208
FT CONFLICT 245 245
SQ SEQUENCE 417 AA; 47366 MW; B1CFF212D830305E CRC64;

```

```

Query Match 42.4%; Score 567; DB 1; Length 417;
Best Local Similarity 41.3%; Pred. No. 1.1e-45;
Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

```

```

OY 1 ASASYEYHSLNEIYSWIEFTEHRPDM.LTKIHSSFEKYP.LYLVKSGKQOTAKNAI 60
D 111 ATGSHTEKINKMETTEHWQVATENDALISRSVIGTTEGRAIVILKV-GKAGQNKPAI 169
OY 61 WIDGCIHAREWISPARCLWFI-----GH-----NRMR 83
D 170 FMDGCFHAREWISPARCLWVFAVRYTEYEAHMEFELNDLFYVLPVINDIGYITWTNR 229
OY 84 NNRMRKNSFYANNHCIGDNLNRFASKHMCCEGASSSCSETYCGLYPEPEVAVASFRLRN 143
D 230 SRPWRKTRSTRHTGSSCIGDPPNRF-DAGWCEIGASRNCDDEYCGPAASEKETALAD 288
OY 144 FLRRNINOKATISMHSYQHIYFPEYSTRSKSKDHEELSLVASEAVRAIEKTSKTRTY 203

```



```

DB 269 PIRNLSSTKALUTHSHSQQMMIYPTSTAYKIGENNAELNALAKATVKEL-ASLHGCTKT 347
QY 204 HGHGSETLYLAPGGDDMYDLGIRYSFT 232
DB 348 YGPGATITVPAAGSGDDMAYDGGIRYSFT 376

RESULT 6
CBPC_MOUSE STANDARD: PRT: 417 AA.
ID CBPC_MOUSE
AC P15089:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Seyaiah W.E.,
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT novel member of the carboxypeptidase gene family."
RL J. Biol. Chem. 264:20094-20099(1989)
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: J05118; AAA37369.1; -.
DR PIR: A34487; A34487.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR MGD: MGI:88479; CPA3.
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_carbopept.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF00246; Zn_carbopept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal.
FT SIGNAL 1
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
SQ SEQUENCE 417 AA; 48790 MW; A2B300A066D1BA6D CRC64;

```

Query Match 41.9%; Score 561; DB 1; Length 417;
 Best Local Similarity 43.7%; Pred. NO. 3.9e-45;
 Matches 115; Conservative 37; Mismatches 71; Indels 40; Gaps 5;

```

QY 6 YEOYHSLNELTYSWIEPTERHPDMLTKIHGSSFEKYPILYLKVSCKEOTAKNAIWDG 65
DB 116 YAKYNDMDIVSWTEKMLKHEPMWSRIKIGSTVEDNPLYLKI-GKKDGERKAIPMDG 174
QY 66 IHAREWISPAFCIMFI-----GHN-----RMWR 88
DB 175 IHAREWISPAFCIMFIYQATKSTGKKNKIMTKLLDRNPFYVLPYFNVDGIWSTQDRMR 234
QY 89 KNRSPYANNHCIGTDLNRNPFASKHCEGASSSCSEYCGLYPESEPEVKAASFRLRN 148
DB 235 KNRSRNQNSTCIGTDLNRNF-DVSMDSPTNKNPCNLNVRGPAPESEKETKATNFIKSH 293
QY 149 INQIKYISMHSYSOHIVPEYSTRKSKDHEELSLVASEAVAIKTSKNRTYTHGHS 208
DB 294 LNSIKRYIFHSYSOMLLPYGTFFKLPNNHDLKVARIAVDAL-STRETRYIYGPIA 352
QY 209 EFLYLAPEGGDDMYDLGIRYSF 231
DB 353 STIYKTSGLMDWYDLGIRKTF 375

RESULT 7
CBPC_RAT STANDARD: PRT: 309 AA.
ID CBPC_RAT
AC P21961:
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE A3).
GN CPA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RX MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., Le Trong H., Woodbury R.G., Walsh K.A.,
RA Neutath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules."
RL Biochemistry 30:648-655(1991).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
DR PIR: A33118; A33118.
DR PIR: A38395; A38395.
DR HSSP: P09955; INSA.
DR MEROPS: M14.010; -.
DR InterPro: IPR000834; Zn_carbopept.
DR Pfam: PF00246; Zn_carbopept. 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; zinc.
FT METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 196 196 ZINC (BY SIMILARITY).
FT ACT_SITE 248 248 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 270 270 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 65 78 BY SIMILARITY.
FT DISULFID 137 160 BY SIMILARITY.
FT VARIANT 1 1 MISSING (IN THE MAJOR FORM).
SQ SEQUENCE 309 AA; 35786 MW; 20330FABC3EB83EF CRC64;

```

Query Match 41.3%; Score 552; DB 1; Length 309;
 Best Local Similarity 42.6%; Pred. NO. 1.8e-44;
 Matches 112; Conservative 39; Mismatches 72; Indels 40; Gaps 5;

GN CPA.
OS Cpa. taurus (Bovine).
OC Eutaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91151335; PubMed=1998496;
RX le Huerou I., Guilloteau P., Touleec R., Pulgserver A., Wicker C.;
RT "Cloning and nucleotide sequence of a bovine pancreatic
RL preprocarboxypeptidase A cDNA.";
RN Biochem. Biophys. Res. Commun. 175:110-116(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=96096549; PubMed=8522204;
RA Goo J.H., Kim K.H., Choi K.Y.;
RT "Cloning, sequencing and expression of the gene encoding a major
RL allotypic preprocarboxypeptidase A from bovine pancreas.";
RN Gene 165:333-334(1995).
RN [3]
RP SEQUENCE OF 111-417.
RX MEDLINE=71155183; PubMed=5102489;
RA Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RN chymotryptic peptides of the cyanogen bromide fragment F.I.";
RL Biochemistry 10:938-950(1971).
RN [4]
RP REVISIONS TO 138 AND 141.
RX MEDLINE=72138789; PubMed=5143102;
RA Petra P.H., Hermanson M.A., Walsh K.A., Neurath H.;
RT "Characterization of bovine carboxypeptidase A (Allan).";
RN Biochemistry 10:4023-4025(1971).
RN [5]
RP SEQUENCE OF 17-120.
RX MEDLINE=89150306; PubMed=3147705;
RA Wade R.D., Haas G.M., Kumar S., Walsh K.A., Neurath H.;
RT "The amino acid sequence of the activation peptide of bovine pro-
RN carboxypeptidase A.";
RL Biochimie 70:1137-1142(1988).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
RX MEDLINE=83294519; PubMed=6887246;
RA Rees D.C., Lewis M., Lipscomb W.N.;
RT "Refined crystal structure of carboxypeptidase A at 1.54-A
RN resolution.";
RL J. Mol. Biol. 168:367-387(1983).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX MEDLINE=96003618; PubMed=7556081;
RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT "The three-dimensional structure of the native ternary complex of
RN bovine pancreatic procarboxypeptidase A with proproteinase E and
RL chymotrypsinogen C.";
RN Embo J. 14:4387-4394(1995).
RN [8]
RP VARIANT ALLELIC.
RX MEDLINE=69283620; PubMed=5817619;
RA Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Identification of the amino acid replacements characterizing the
RN allotypic forms of bovine carboxypeptidase A.";
RL Biochemistry 8:2762-2768(1969).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER. THE ZMOGEN IS SECRETED AS A TERNARY COMPLEX
CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC PROTEINASE E.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -1- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M61851; AAA30426.1; -
DR EMBL; M61851; AAA30427.1; -
DR EMBL; Z33906; CAA83955.1; -
DR PIR; JN0126; CPB0A.
DR PIR; A31406; A31406.
DR PDB; 3CPA; 15-JAN-87.
DR PDB; 4CPA; 22-OCT-84.
DR PDB; 5CPA; 15-JAN-87.
DR PDB; 6CPA; 15-OCT-91.
DR PDB; 7CPA; 31-JAN-94.
DR PDB; 8CPA; 31-JAN-94.
DR PDB; 1CBX; 31-JAN-94.
DR PDB; 1CPS; 15-OCT-94.
DR PDB; 2CTB; 31-JAN-94.
DR PDB; 2CRC; 31-JAN-94.
DR PDB; 1ARL; 01-AUG-96.
DR PDB; 1ARM; 17-AUG-96.
DR PDB; 1BAV; 01-APR-97.
DR PDB; 1YME; 12-FEB-97.
DR PDB; 1CPX; 05-AUG-98.
DR PDB; 1PYT; 27-JAN-97.
DR MEROPS; M14.001; -
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept. 1.
DR PRINTS; PR00765; CROXYPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; zinc; zymogen; signal;
KW 3d-structure; Polymorphism.
FT SIGNAL 1 16
FT PROPEP 17 110
FT CHAIN 111 419
FT METAL 179 179
FT METAL 182 182
FT METAL 306 306
FT ACT_SITE 358 358
FT ACT_SITE 380 380
FT DISULFID 248 271
FT VARIANT 289 289
FT VARIANT 338 338
FT VARIANT 415 415
FT CONFLICT 95 95
FT CONFLICT 199 199
FT CONFLICT 203 203
FT CONFLICT 224 224
FT CONFLICT 232 232
FT CONFLICT 295 295
FT CONFLICT 114 116
FT TURN 119 120
FT TURN 125 138
FT HELIX 140 142
FT STRAND 143 150
FT STRAND 152 153
FT STRAND 156 162
FT STRAND 171 176
FT TURN 180 181
FT HELIX 183 199
FT TURN 200 202
FT HELIX 204 212
FT STRAND 214 218
FT HELIX 223 231
FT TURN 232 232
FT TURN 234 235
I -> V (IN ALLELIC VARIANT).
E -> A (IN ALLELIC VARIANT).
L -> V (IN ALLELIC VARIANT).
S -> L (IN REF. 5).
D -> N (IN REF. 3).
D -> N (IN REF. 3).
D -> N (IN REF. 3).
O -> E (IN REF. 3).
D -> N (IN REF. 3).
ACTIVATION PEPTIDE.
CARBOXYPEPTIDASE A.
ZINC.
ZINC.
ZINC.
PROTON DONOR.
NUCLEOPHILE.

```

RP SEQUENCE FROM N.A. PubMed=3182872:
R3 MEDLINE=69034324:
RA Clauser E., Gardelli S.J., Craik C.S., Macdonald R.J., Rutter W.J.,
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family."
RT J. Biol. Chem. 263:17837-17845(1988).
CC -1 CATALYTIC ACTIVITY: peptidyl-L-amino acid + H2O = peptide + L-
CC amino acid.
CC -1 SUBUNIT: MONOMER.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL: V01232; CAA2542.1; -
CC EMBL: J00713; AAA40893.1; -
CC EMBL: M23990; AAA40955.1; -

```

DR	EMBL	M23985	AAAA0355.1	JOINED.
DR	EMBL	M23986	AAAA0355.1	JOINED.
DR	EMBL	M23987	AAAA0355.1	JOINED.
DR	EMBL	M23988	AAAA0355.1	JOINED.

DR PIR; A00911; CRITA.
DR PIR; B32129; B32129.
DR HSSP; P00730; LEVT.

DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR ProDom; PD00015

```
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.  
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.  
DR Hydrolase; Carboxypeptidase; Metalloprotease; M1-2; E=0.000001
```

FT	CHAIN	17	110	ACTIVATION PEPTIDE.
FT	CHAIN	111	419	CARBOXYPEPTIDASE AI.
TT	METAL	179	179	ZINC (BY SIMILARITY)

FT	ACT_SITE	358	ZINC (BY SIMILARITY).
FT	ACT_SITE	358	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	380	NUCLEOPHILE (BY SIMILARITY).

FT CONFLICT	261	263	FGM -> LGM (IN REF. 1).
FT CONFLICT	347	347	K -> E (IN REF. 1).
FT CONFLICT	347	347	

Query Match	Score	DB 1;	Length
Best Local Similarity	36.5%	488.5;	419;
	30.4%	488.5;	419;

6 YEQYHSLNEIYSWTEFITERHDM LTKIHIGSSFEKYPPLYVKVSGKEQTAKNAIWI DCG 65

```

0Y      66 IHAREWISPAFLMFIH-----NMR 88
|::|::|:|::|

```

```
89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSCSETTCGLYPESEPEVKAVASFLRRN 148
```

OY 149 INQIKAYISMHSQIIVPYSTRKSKDHELSIVASEAVAIKTSKNTYTHGHS 208
 DB 297 GN-IFKAFISHSQILLYGYTSEPAPQDELQDLAKSAVATL-TSLHGTFRKYSII 354
 OY 209 EFLYLAPGGDDMIYDLGIRKYSFT 232
 DB 355 DRIYQASGSIIDMTYISQGIKYSFT 378

RESULT 11
 CBP2_RAT STANDARD: PRT; 417 AA.
 AC P19222;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
 GN CPA2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=89034323; PubMed=3182871;
 RA Gargiell S.J., Craik C.S., Clausner E., Goldsmith E.J., Stewart C.-B.,
 RA Graf M., Rutter W.J.;
 RT "A novel rat carboxypeptidase, CPA2: characterization, molecular
 RT cloning, and evolutionary implications on substrate specificity in
 RT the carboxypeptidase gene family.";
 RL J. Biol. Chem. 265:17828-17836(1990).
 RN [2]
 RP SEQUENCE OF 131-143 FROM N.A.
 RX MEDLINE=95386501; PubMed=7657630;
 RA Normant E., Gros C., Schwartz J.C.;
 RT "Carboxypeptidase A isoforms produced by distinct genes or
 RT alternative splicing in brain and other extrapancreatic tissues.";
 RL J. Biol. Chem. 270:20543-20549(1995).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=92105124; PubMed=1761558;
 RA Fleming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
 RT "Structural evolution of an enzyme specificity. The structure of rat
 RT carboxypeptidase A2 at 1.9-A resolution.";
 RL J. Biol. Chem. 266:24606-24612(1991).
 CC -1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
 CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
 CC residues.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
 CC ZINC CARBOXYPEPTIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M23721; AAA40956.1; -
 DR EMBL; M23714; AAA40956.1; JOINED.
 DR EMBL; M23715; AAA40956.1; JOINED.
 DR EMBL; M23716; AAA40956.1; JOINED.
 DR EMBL; M23717; AAA40956.1; JOINED.
 DR EMBL; M23718; AAA40956.1; JOINED.
 DR EMBL; M23719; AAA40956.1; JOINED.
 DR EMBL; M23720; AAA40956.1; JOINED.
 DR EMBL; S79837; -; NOT_ANNOTATED_CDS.
 DR PIR; A32188; A32188.
 DR HSSP; P48052; IAYE.
 DR MEROPS; M14.002; -.
 DR InterPro; IPR003146; Propep_M14.

DR InterPro; IPR000834; Zn_Carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_Carboxypept; 1.
 DR PRINTS; PR00765; CARBOXYPTASEA.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 DR PROSITE; PS00134; CARBOXYPEPT_ZN_3; 1.
 KW Hydrolyase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL.
 FT PROPEP 1 112
 FT CHAIN 113 417
 FT METAL 177 177
 FT METAL 180 180
 FT METAL 304 304
 FT METAL 378 378
 FT ACT_SITE 246 269
 FT DISULFID 318 352
 FT DISULFID 318 352
 SO SEQUENCE 417 AA; 46912 MW; BEDCC41A830F2D45 CRC64;

Query Match 36.1%; Score 483; DB 1; Length 417;
 Best Local Similarity 38.2%; Pred. No. 8e-38;
 Matches 109; Conservative 47; Mismatches 75; Indels 54; Gaps 9;

OY 6 YEOYHSINLEIYSNIEPTTERHPMLTKIHIGSSPEKPLVYKVS-GKEQTAKNAIWDIC 64
 DB 118 FEAYHTLEIYQENDNLVAENPGLVSKVNLGSSFENPMNVLFSTGDD--KPAIWLDA 174
 OY 65 GHAREWISPAFCILM-----FT-----GH-----NRMW 87
 DB 175 GHAREWYOTATMTANKIASDYGTDPALITSLNTIDIFLPTNDGIVFSQTTNRW 234
 OY 88 RKNRSFYANNHCIGTDLNRNFRASKHWCESGASSCSSEYCGLYPESEPEVKAASFLRR 147
 DB 235 RKRFRSKRSGSCGVDPENRNM-DANFGPGASSPSCSDSYHGPKPNEVEVKSIYDPIKS 293
 OY 148 NINQIKAYISMHSQIIVPYSTRKSKDHELSIVASEAVAIKTSKNTYTHGHS 207
 DB 294 H-GVKKAFITLHYSQILMPYGYKCRKPDDELVAQKAQAL-KRLHGTSTYKVGPI 351
 OY 208 SETLYLAPGGDDMIYDLGIRKYSFTSN-----PVEKLLP 242
 DB 352 CSVIYQASGSIIDMTYISQGIKYSFTAFELRLRTAFYGLLPKQILP 396

RESULT 12
 CBP1_HUMAN STANDARD: PRT; 419 AA.
 AC P15085;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
 GN CPA1 OR CPA.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=93038569; PubMed=1417781;
 RA Catasus L., Villegas V., Pascual R., Aviles F.X.,
 RA Wicker-Planquart C., Pulgiver A.;
 RT "cDNA cloning and sequence analysis of human pancreatic
 RT procarboxypeptidase A1.";
 RL Biochem. J. 287:299-303(1992).
 RN [2]
 RP SEQUENCE OF 17-42.
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT "Purification and properties of five different forms of human
 RT procarboxypeptidases.";
 RL Eur. J. Biochem. 179:609-616(1989).

```

CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O -> peptide + L-
CC amino acid.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X67318; CAA47732.1; -.
DR PIR; S02810; S02810.
DR PIR; S29127; S29127.
DR HSSP; P00730; 1PYT.
DR MEROPS; M14.001; -.
DR MIM; 114850; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CROBOXPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR HydroLase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 182 182 ZINC (BY SIMILARITY).
FT METAL 306 306 ZINC (BY SIMILARITY).
FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 248 271 BY SIMILARITY.
SO SEQUENCE 419 AA; 47140 MW; 439FAFFREB58B1 CRC64;

Query Match 34.8%; Score 465.5; DB 1; Length 419;
Best Local Similarity 37.9%; Pred. No. 3.5e-36;
Matches 100; Conservative 44; Mismatches 79; Indels 41; Gaps 5;

DQ 6 YEOYHSLSNEIYSWIEFTEHRPDMLTRKIHGSSFEKYLVLKYSCKQTKAKNAIWDG 65
DQ 119 YATHTLEETIDYDFDLVAENPHLVSKIQENTYEGRPYLVKFS-TGSGSKRPAMIDTG 177
DQ 66 IHAREWISPAFLCWF-----GHNRMWR 88
DQ 178 IHSREWYTOAGGVFAKKIINDYGDAAFTAILDTLDFLEIVNPDGFATHTSTNRMR 237
DQ 89 KNSRFYANNHCIGTDLRN--FASKHMCCEGASSSCSEYTCGLYPESEBEVKAASFLRN 148
DQ 238 KTRSHYTGSLICIGDVPNNMADAGFL-SGASSNPSCSEYTHGKFPANSEVEKSIYDFVDH 296
DQ 149 INOKAIVSMHSYSOHIVPEPYSTRSKSDHELSIVASEAVRAIEKTSKTRTHGHS 208
DQ 297 GN-ITKAISHISQSLMLPYGVTPEVPDDELDSLKAVALT-ASLVGCTKFNYSII 354
DQ 209 ETLVLAPGGDDWITDGIKYSFT 232
DQ 355 KAIYQASGSTIDWITYSOGIKYSFT 378

RESULT 13
CBP4_HUMAN STANDARD; PRT; 421 AA.
AC G0U142;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).

```

```

GN CPA4 OR CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99310537; PubMed=10383164;
RA Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT deacetylase inhibitors during differentiation of prostate epithelial
RT cancer cells."
RL Cancer Res. 59:2981-2988(1999).
CC -1- FUNCTION: Could be involved in the histone hyperacetylation
CC pathway.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF095719; AAF23230.1; -.
DR HSSP; P48052; IAYE.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CROBOXPTASEA.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR HydroLase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 113 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 114 421 CARBOXYPEPTIDASE A4.
FT METAL 181 181 ZINC (BY SIMILARITY).
FT METAL 184 184 ZINC (BY SIMILARITY).
FT METAL 308 308 ZINC (BY SIMILARITY).
FT ACT_SITE 382 382 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 250 273 BY SIMILARITY.
SO SEQUENCE 421 AA; 47379 MW; 91883D5B705BC6A CRC64;

Query Match 34.2%; Score 458; DB 1; Length 421;
Best Local Similarity 38.4%; Pred. No. 1.8e-35;
Matches 103; Conservative 35; Mismatches 82; Indels 48; Gaps 6;

DQ 6 YEOYHSLSNEIYSWIEFTEHRPDMLTRKIHGSSFEKYLVLKYSCKQTKAKNAIWDG 65
DQ 120 YGAYHSLSLEATYHEMDNTLAADPDLARVYKIGHSPENPMYLVKSTGKYRRAPVWLNAG 179
DQ 66 IHAREWISPAFLCWF-----FT-----GHNRMWR 88
DQ 180 IHSREWISQATVAITWARTAKIYSDYQORDPATISILEKMDIFILPVANPDGYVYTGQONLMR 239
DQ 89 KNSRFYANNHCIGTDLRN----FASKHMCCEGASSSCSEYTCGLYPESEBEVKAASFLRN 144
DQ 240 KTRSRNPGSSCIGADPPKRNMMASPAK-----GASDPCESEVYHGPIANSEVEKSYVDF 294
DQ 145 LRRNINQIKAYISMHSYSOHIVPEPYSTRSKSDHELSIVASEAVRAIEKTSKTRTHGHS 204
DQ 295 IOKHGN-FKGFIDHYSQSLMLPYGYSVKKAPDAEELDVAAALAKALASVS-GTEIYQV 352
DQ 205 GHGSETIYLPAGGDDWITDGIKYSFT 232
DQ 353 GPCTTVYPAAGSSIDWAYDNGIKFAFT 380

RESULT 14

```

CPB2_HUMAN
ID CPB2_HUMAN STANDARD; PRT; 417 AA.
AC P48052;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
RC TISSUE=Pancreas;
RA MEDLINE=95204457; PubMed=7896805;
RA Catusus L., Vendrell J., Aviles F.X., Carreira S., Pulgarer A.,
RA Billeter M.;
RT "The sequence and conformation of human pancreatic
RT procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT three-dimensional model.";
RL J. Biol. Chem. 270:6651-6657(1995).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98111000; PubMed=9450539;
RA Reverter D., Garcia-Saez I., Catusus L., Vendrell J., Coll M.,
RA Aviles F.X.;
RT "Characterisation and preliminary X-ray diffraction analysis of human
RT pancreatic procarboxypeptidase A2.";
RL FEBS Lett. 420:7-10(1997).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98046021; PubMed=9384570;
RA Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT "The three-dimensional structure of human procarboxypeptidase A2.
RT Deciphering the basis of the inhibition, activation and intrinsic
RT activity of the zymogen.";
RL EMBO J. 16:6906-6913(1997).
CC -1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC residues.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14, ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U19977; AAA74425.1; -
DR PDB: 1AYE; 13-JAN-99.
DR MEROPS: M14.002; -
DR MIM: 600688; -
DR InterPro: IPR003146; Propep_M14.
DR InterPro: IPR000834; Zn_Carboxypeptidase.
DR Pfam: PF02244; Propep_M14; 1.
DR Pfam: PF02246; Zn_Carboxypeptidase; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW 3D-structure.
FT SIGNAL 1 16 POTENTIAL.
FT PROPEP 17 112 ACTIVATION PEPTIDE.
FT CHAIN 113 417 CARBOXYPEPTIDASE A2.
FT METAL 177 177 ZINC.
FT METAL 180 180 ZINC.
FT METAL 304 304 ZINC.
FT ACT_SITE 378 378 NUCLEOPHILE.
FT DISULFID 246 269

FT DISULFID 318 352
SQ SEQUENCE 417 AA; 46840 MW; 608107330A2DBC58 CRC64;
Query Match 34.1%; Score 456; DB 1; Length 417;
Best Local Similarity 36.5%; Pred. No. 2.7e-35;
Matches 103; Conservative 45; Mismatches 80; Indels 54; Gaps 7;
QY 9 YHSLNEIYSWIEFTTERHPDMLTKRHIGSSFEYRPLVLKVS-GKEQTAKNAIWDGIR 67
DQ 121 YHRLLEISQEMDLVAHEHPLVSKVNISSFEHNPVNLKFSIGD--KPAIWLDAIGH 177
QY 68 ARWISPAFLMFTG-----HNRMRKN 90
DQ 178 ARWVTOATALMANKIVSDYGRDPSITSLDALDIFLPTVPNDGYVFSQTKRNRMRKT 237
QY 91 RSFYANNHCGTDLNRFASKHWCBSGSSSEYFCGLYPPSEPVKAVASFLRNIN 150
DQ 238 RSKVSGSLCVDPNRRW-DAGFGGPGASSNPSDSTHGFSAENSEVKSIVDFIKSH-G 295
QY 151 QIKAYISMHSQHIYEPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGSET 210
DQ 296 KVKAFILHSYQLMFPYCYKTKTDDPELSEVAKAQS-RLHGTIKYVGPICSV 354
QY 211 LYLAPGGDDIVDLGKISFTSN-----PVEKILP 242
DQ 355 IYQASGSDISYDYGKYSFAFELRDTGRYGFLLPARQILP 396
RESULT 15
CPB2_SIMV1
ID CPB2_SIMV1 STANDARD; PRT; 304 AA.
AC P42788;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.-) (fragment).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gut;
RX MEDLINE=94093864; PubMed=8269093;
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly *Simulium vittatum* encoding
RT trypsin-like and carboxypeptidase-like proteins.";
RL Insect Mol. Biol. 1:149-163(1993).
CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14, ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L08481; AAA18531.1; -
DR HSP: P48052; 1AYE.
DR InterPro: IPR000834; Zn_Carboxypeptidase.
DR Pfam: PF02246; Zn_Carboxypeptidase; 1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 58 58 ZINC (BY SIMILARITY).
FT METAL 61 61 ZINC (BY SIMILARITY).
FT METAL 61 61

OM of: US-09-980-881-4 to: N_Geneseq_032802: * out_format : pfs
Date: Sep 18, 2002 7:03 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ .p2n.model -DEV=x1h
-Q=/cgn2.1/USPRO.spool/US09980881/runat_16092002_140130_10714/app_query.fasta.1.1126
-DB=N_Geneseq_032802 -QMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOAPEXT=0.500
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum2
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09980881.ecgn1.1.0 -NCPU=6 -ICPU=3 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-4
Query length: 246
Database: N_Geneseq_032802: *
Database sequences: 1736436
Database length: 858457221
Search time (sec): 386.310000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC1962 +	1325.00	2699.78	5.0e-142	15	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAV74302 +	1272.00	2592.93	4.4e-136	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAV11671 +	1272.00	2589.27	7.1e-136	17	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2846 +	1272.00	2589.27	7.1e-136	17	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:AAQ41001 +	1267.00	2578.96	2.7e-135	17	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAF18005 +	902.00	1828.84	1.6e-93	14	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT5760 +	610.00	1231.43	3.0e-60	92	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ50601 +	582.00	1173.77	5.0e-57	92	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ50600 +	582.00	1170.58	7.4e-57	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42500 +	567.00	1141.30	2.9e-55	99	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT2497 +	567.00	1141.30	3.2e-55	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT4294 +	567.00	1139.21	4.2e-55	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT1795 +	567.00	1139.21	4.2e-55	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42506 +	567.00	1139.02	4.3e-55	12	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAAT4804 +	560.00	1126.86	2.0e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT42511 +	560.00	1126.80	2.0e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2787 +	560.00	1126.80	2.0e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AAAT42512 +	559.00	1124.73	2.7e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2788 +	559.00	1124.73	2.7e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAD03837 +	556.50	1117.13	7.1e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2806 +	556.00	1118.55	5.9e-54	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2790 +	554.00	1114.42	1.0e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2801 +	554.00	1114.42	1.0e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2789 +	553.00	1112.36	1.3e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2791 +	552.00	1110.30	1.7e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2805 +	551.00	1108.24	2.2e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2803 +	551.00	1108.24	2.2e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2804 +	551.00	1108.24	2.2e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6477 +	550.50	1106.51	2.8e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6479 +	550.50	1106.51	2.8e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6476 +	550.50	1104.48	3.6e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6478 +	550.50	1104.48	3.6e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAT6478 +	550.50	1100.94	5.6e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2808 +	549.00	1104.11	3.8e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2800 +	549.00	1104.11	3.8e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT2800 +	549.00	1097.58	8.7e-53	18	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAAT1819 +	549.00	1097.58	8.7e-53	18	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT1731 +	549.00	1095.95	1.1e-53	21	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT6802 +	548.00	1102.05	4.9e-53	10	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:AAAT6807 +	546.00	1097.93	8.3e-53	10	

/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAD03839 +	525.50	1047.63	5.3e-50	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAD05477 +	492.00	985.13	1.6e-46	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAD03838 +	471.50	944.40	3.0e-44	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC98830 +	465.50	929.98	1.9e-43	
/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAAT46119 +	465.50	929.95	1.9e-43	

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAAC1962

seq_documentation_block:

ID AAC1962 standard; cDNA; 1573 BP.

XX AAC1962;

XX 01-MAR-2001 (first entry)

XX Human brain carboxypeptidase B cDNA.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta amyloid precursor protein; vaccine;
XX cerebrotrophic; antialzheimer's; neurotrophic; neuroprotective;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 18..1100 /tag= a /product= "carboxypeptidase B"

XX WO20006717-A1.

XX 09-NOV-2000.

XX 01-MAY-2000; 2000WO-JP02878.

XX 30-APR-1999; 99JP-0125169.

XX (MATS/) MATSUMOTO A.

XX Matsumoto A;

XX WPI: 2000-687534/67.

XX P-PSDB: AAB11457.

XX Human brain carboxypeptidase B isolated from the hippocampus useful for
XX screening agents for the treatment of Alzheimer's and other brain
XX disorders -

XX Claim 2b; Page 64-68; 84pp; Japanese.

XX This invention describes a novel protein with peptidase activity
XX against brain beta-amyloid precursor protein which has been isolated from
XX human hippocampus and which has cerebrotrophic, antialzheimer's,
XX neurotrophic, neuroprotective and hemostatic activity and which can be used
XX as a vaccine or for gene therapy. The protein, and compounds identified
XX by screening as promoters or inhibitors of its activity, are used to
XX regulate beta-amyloid accumulation in the brain and treat or prevent
XX diseases in which this occurs, such as Alzheimer's, senile dementia,
XX inherited cerebral hemorrhage, Down's syndrome, and head trauma.

XX Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;

XX alignment_scores:

Quality: 1325.00	Length: 246
Ratio: 5.408	Gaps: 0
Percent Similarity: 99.593	Percent Identity: 98.780

XX alignment_block:

XX US-09-980-881-4 x AAC1962 ..
XX Align seg 1/1 to: AAC1962 from: 1 to: 1573


```

83 ..... 83
593 TAACATCAATTCATAGGATAATAGGCAATATACCAATCTCTGAGGCTT 642
83 ..... 83
643 GTGGATTTCATGTATGCCGGTGTAAATGTGGACGGTTATAGCTACTC 652
84 ..... AsnArgMetIlePArgLysAsnArgSerPheTyrAlaAsn 97
693 ATGGAAAAAGAAATCGAATGTGGAGAAAGAACCTTCTTCTATAGCGACA 742
97 snHicSylIegLyrAspLeuAsnArgAsnPheAlaSerLysSHSTP 113
743 ATCATGTGATCGGACAGACCTGAAATAGGAATCTTCTTCCAAACCTGG 792
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
793 TGTGAGGAAGGTGCATCCAGTTCTCATGCTCGAAACCTACTGTGACT 842
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
843 TTATCTGTAGTCAGACAGACAGAACTGAAGGAGCTAGTTCTTGAGAA 892
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
893 GAAATATCAACCGATTAAAGCATATCATCAGCATTCATCATCTCCAG 942
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHis 180
943 CATATAGTGTTCATATTCCTATACACGAAATAAAGCAAAACCATGTA 992
180 uGluSerLeuValAlaSerGluAlaValAlaArgAlaIleGlyLysThr 197
993 GGAACGTCTCTAGTAGCCAGTAGACAGTCTGCTATGTGAAAAACTA 1042
197 erLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
1043 GTMAAAATACAGGTATACACATGCGCATGCGTCAAGAAACCTTATCTTA 1092
214 AlaProGluGlyLysAspArgPheTyrAspLeuGlyLysTyrSer 230
1093 GCTCTGGAGGTGGGACATGTGATCTATGATTTGGCATCAAAATATTC 1142
230 r.Phe..... 231
1143 GTTACAATTCGACTCGAGATACGGGACATACGATTTCTGTGCGCG 1192
232 .....ThrSerAsnProProValGluLysLeuLeuProLeuSerLeu 246
1193 AGCGTTACATCAAAACCCACTGTAGAGAAAGCTTTGCGCTGTCTAA 1242
246 s 246
1243 A 1243
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:AA11671
seq_documentation_block:
ID AA11671 standard; DNA; 1749 BP.
AC AA11671;
AT 12-APR-1996 (first entry)
DE Human plasma carboxypeptidase B coding sequence.
KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
purification; plasminogen; affinity column; ss.
XX Homo sapiens.
XX
XX key Location/Qualifiers

```

```

FT CDS 41..1312
FT FT /*tag= a
FT FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT FT /*tag= b
FT mat_peptide 107..1309
FT FT /*tag= c
XX
XX US5474901-A.
XX
XX 12-DEC-1995.
XX
XX 01-FEB-1991; 9105-0649591.
XX PF
XX 01-FEB-1991; 9105-0649591.
XX PR
XX 14-OCT-1992; 9205-0959944.
XX PR
XX 15-DEC-1993; 9305-0167727.
XX PR
XX 19-JUL-1994; 9405-0277540.
XX
XX (GENTECH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI: 1996-039508/04.
XX DR
XX P-PSDB; AAR90293.
XX
XX Antibody to human plasma carboxypeptidase B - useful for detecting
XX PT and purifying hPCPB for use in treating clotting disorders e.g.
XX PT haemophilia A
XX
XX PS Disclosure; Figure 4; 40pp; English.
XX
XX An antibody which specifically binds human plasma carboxypeptidase B
XX CC (hPCPB) and does not cross react with other carboxypeptidases is
XX CC useful for the detection of hPCPB in vitro. The antibody is also
XX CC used for purifying hPCPB from a sample. Purification comprises
XX CC passing a sample thought to contain hPCPB over either a column to
XX CC which antibody has been bound, or a plasminogen affinity column,
XX CC eluting the column and then recovering the fraction containing the
XX CC hPCPB.
XX
XX SQ Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

alignment_scores:
Quality: 1272.00 Length: 301
Ratio: 5.171 Gaps: 3
Percent Similarity: 81.728 Percent Identity: 81.395

alignment_block:
US-09-980-881-4 x AA11671 ..
Align seg 1/1 to: AA11671 from: 1 to: 1749

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSer 17
383 GCCGCCGCATCGTACTATGAACAGTACACTCACTAATGAATCAATCTATTC 432
17 rTTPriIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
433 TTGGATAGAAATTTAACTGAGAGCAATCCGATATGCTTAACAAAATGCC 482
34 IsIleGlySerPheGluLysTyrProLeuTyrValIleLysValSer 50
483 ACATGTGATCTCTCATTTGAGAGATACCACTCTATGTTTAAAGGTTCT 532
51 GlyLysGluGlnThrAlaLysAsnAlaIleTTPriLeaSPcysGlyIleH 67
533 GGAAGAAACAAACAGCCAAATGATGATGATGATGATGATGATGATGATGAT 582
67 sAlaArgGluTTPriLeaSerProAlaPheCysLeuTTPriPheIleGlyHis 83
583 TGGCAGAGAAATGATCTCTCTGCTTCTGCTTGTGTGTGATGATGATGATGAT 632

```

```

83 ..... 83
633 TAACCAATTCATGGGATATAGGCAATATACCAATCTCTGAGGCTT 682
83 ..... 83
683 GTGGATTTCTATGTTATGCCGGTGTAAATGTGGAGCGTTATGACTACTC 732
84 .....AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnA 97
733 ATGAAAAAAGAAATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAACA 782
97 snHIScysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrp 113
783 ATCATTTGCATCGGACAGACCTGAATAGAGAACTTCTCTCAACACCTCG 832
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
833 TGTGAGAGAGGTGCATCCAGTCTCTCATCTCGAAACCTACTGTGGACT 882
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgA 147
883 TTATCTGTGATCAGAACCAAGTAGAGGCACTGGCTACTTCTTGAGAA 932
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
933 GAAATATCAACCAAGATTAAGCATATCATCATCATCATCATCATCATC 982
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAsnHisG 180
993 CATATAGGTTTCCATATTCCTATACAGCAAGTAAAGCAAGCAAGCAAGA 1032
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysHis 197
1033 GGAATGCTCTAGTAGCCAGTAGAAGCAGTCTGCTCTATTGGAAGAACTA 1082
197 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
1083 GTAAATAATACCGGTATACACATGGCCATGGCTCAGAAACCTTAACCTTA 1132
214 AlaProGlyGlyLysAspAspTrpIleTyrAspLeuGlyLysTyrSer 230
1133 GCTCTGGAGGTGGGAGCATGTGATCTATGATTGGGCAATCAATATATTC 1182
230 r.Phe..... 231
1183 GTTATCAATTGAACTTCGAGATACGGGCAATACGATTCCTGTGCCGG 1232
232 .....ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuL 246
1233 AGCGTTACATCAAAACCCACCTGTAGAGAAAGCTTTGGCCGCTGTCTTAA 1282
246 s 246
1283 A 1283
seq_name: /STDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:AAAT62846
seq_documentation_block:
ID AAAT62846 standard; DNA: 1749 BP.
XX
AC AAAT62846;
XX
DT 08-MAY-1997 (first entry)
XX
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Human: plasma carboxypeptidase B; PCPB; haemostatic regulation;
XX plasma; plasminogen; ss.
XX
OS Homo sapiens.
XX

```

```

FH Key Location/Qualifiers
FT CDS 41..1312
FT .. /*tag= a
FT .. /product= Human PCPB
FT sig_peptide 41..106
FT .. /*tag= b
FT mat_peptide 107..1309
FT .. /*tag= c
FT misc_binding 134..177
FT .. /*tag= d
FT .. /bound_moiety= 46_bp_probe
XX
XX US5593674-A.
XX
XX 14-JAN-1997.
XX
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 14-OCT-1992; 92US-0959944.
XX
XX 15-DEC-1993; 93US-0167727.
XX
XX 19-JUL-1994; 94US-0277540.
XX
XX 27-APR-1995; 95US-0430787.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI: 1997-099413/09.
XX
XX P-PSDB; AAM14733.
XX
XX Using human plasma carboxypeptidase B in blood coagulation - is
XX functionally related to carboxypeptidase A and pancreas
XX carboxypeptidase B
XX
XX Example 2: Column 37-42; 39pp; English.
XX
XX This sequence encodes human plasma carboxypeptidase B (PCPB) which
XX has a molecular weight under non-reducing SDS-PAGE of approx. 60 kd.
XX PCPB may be used therapeutically in haemostatic regulation. PCPB is
XX purified from human plasma or by transformed cell culture by
XX extraction using plasminogen bound to a solid phase.
XX
XX Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;
XX
XX
XX Alignment_scores:
XX Quality: 1272.00 Length: 301
XX Ratio: 5.171 Gaps: 3
XX Percent Similarity: 81.728 Percent Identity: 81.395
XX
XX alignment_block:
XX US-09-980-881-4 x AAAT62846 ..
XX
XX Align seg 1/1 to: AAAT62846 from: 1 to: 1749
XX
XX 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
XX |||||
XX 383 GCCTCCGCAATCGTACTATGAAACAGTATACACTAATGAAATCTATTTC 432
XX |||||
XX 17 rTrpIleGluPheIleThrGluAlaGlnHisProAspMetLeuThrLysIleH 34
XX |||||
XX 433 TTGGATGAAATTATTAACGTAGAGAGCATCCGATATGCTTACAAAAATCC 482
XX |||||
XX 34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
XX |||||
XX 483 ACATTGGATCTCTCATTTGAGAGAGTACCCACTCTATGTTTAAAGGTTTCT 532
XX |||||
XX 51 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
XX |||||
XX 533 GGAAGAGAACAAACAGCCAAAGATGCAATGATGATGATGATGATGCAATCCA 582
XX |||||
XX 67 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 83

```



```

|||||
583 TGGCAGAGATGATCTCTGCTTCTGCTTGTGCTCATAGGCCATA 632
83 ..... 83
633 TAATCATTTATATGGGATATAGGCAATATACCAATCTCCGAGGCTT 682
83 ..... 83
683 GTGGATTTCTATGATGCCGGTCTTAATGTGACGCTTATGACTACTC 732
84 .....AsnArgMetIrpArgLysAsnArgSerPheTyrAlaAsnA 97
733 ATGAAAAAAGATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAACA 782
97 snHsCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHsItrp 113
783 ATCAATGTCATCGGAACAGACCTGTAATAGACTTGTCTTCCAACACTGG 832
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
833 TGTGAGGAAGGTGCATGCAGTTCCTCATGCTCGAAACCTACTGTGGACT 882
130 UTYrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
883 TTATCTGTAGTCAAGAACGAGAGTGAAGGCACTGGCTACTTCTTGACAA 932
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
933 GAAATATCAACCAAGATTAAGCATATCAGCATCATTCATATACATCCAG 982
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 180
933 CATATAGTGTTCATATATCCATATACGGAAGTAAAGCAAGACCATCA 1032
180 uGluLeuSerLeuValAlaSerGluValArgAlaIleGluLysThr 197
1033 GGAATGCTCTAGTACAGCAGTGAAGCACTTGTCTATTTGAAATACTA 1082
197 erLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
1083 GTAAAAAACCGGTATACACATGCGCATGGCTCAGAAACCTTATATACCTA 1132
214 AlaProGlyGlyLysAspAspTyrIleTyrAspLeuGlyLysTyrSe 230
1133 GCTCTGAGAGGTGGGAGCATTTGATCATATTTGGCATCAAAATATTC 1182
230 r.Phe..... 231
1183 GTTTCACATTTGAACTTCAGATATACGGGACATACGAGATTCTTGCGCG 1232
232 .....ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuL 246
1233 AGCGTTACATCAAAACCCAGCTGTAGAGAAAGCTTTTCCGCTGTCTTAA 1282
246 s 246
1283 A 1283
seq.name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAF18005
seq_documentation_block:
ID AAF18005 standard; DNA; 1400 BP.
XX
AC AAF18005;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
cardioactive; immunomodulatory; muscular active; vulnerary;
gastrointestinal; nephrotropic; antinfective; gynecological;

```

```

KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
PN W020005180-A2.
XX
21-SEP-2000.
XX
08-MAR-2000; 2000WO-US05918.
XX
12-MAR-1999; 99US-0124270.
XX
(HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM.
XX
WPI: 2000-587514/55.
DR P-PSDB; AAB58129.
XX
Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer.
XX
PS Claim 1; Page 507; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences. Their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
XX
alignment_scores:
Quality: 902.00 Length: 238
Ratio: 4.983 Gaps: 4
Percent Similarity: 76.050 Percent Identity: 75.210
alignment_block:
US-09-980-881-4 x AAF18005 ..
Align seg 1/1 to: AAF18005 from: 1 to: 1400
65 GYLIEHLSAIAArgLUTrPILeserProAlaPheCysLeuTrpPheI 81
|||||
10 GAATTCATGCCAGANAAATGATCTCTGCTTCTTCTGCTTGTTCAT 59
|||||
81 eGLYHIS..... 83
60 AGGCATATATACTCAATTCATGGATAATAGGCAATATACCAATCTCC 109
|||||
83 ..... 83
110 TGAGCTTGTGATTTCTATGCTTATGCCGGTCTTAATGTGAGATCTTAT 159
84 .....Asn.ArgMetIrpArgLysAsnArgSerPheT 94
::: |||||||||

```

```

160 GAACTACTCATGGAAGAAAGATGATGTGAGAAAGAACCGTTCTTCT 209
94 Yr1AaSnAaSnHsCysIleGlyThrAspLeuSnArGAsnAraSeraSer 110
|||||
210 ATGCAATCATATTCATTCGATCGAAGACACCTGAAATAGAACTTGTCTCC 259
111 LysHsTrpCysGluGluGlyAlaSerSerSerSerSerSerSerGluTrpTy 127
|||||
260 AAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCAATGCTCGGAACCTA 309
127 rCysGlyLeuTrpProGluSerGluProGluValAlaValAlaSerP 144
|||||
310 CTGTGAGCTTATCTCTAGTCAGAACAGAAAGTGAAGCGCTGCTAGTT 359
144 heLeuArGArGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
|||||
360 TCTTGAAGAAATATCAACCCAGATTAAGCATATCATCAGCATCATTTCA 409
161 TyrSerGlnHisIleValAlaPheProTyrSerTyrThrArgSerLysSerLy 177
|||||
410 TACTCCGAGATATAGTGTTCATATCTTAAACAGAAAGTAAAGCAA 459
177 sAsPHisGluIleuSerLeuValAlaSerGluAlaValAlaGluAlaIleG 194
|||||
460 AGACCATGAGGAAGTGTCTCTAGTACGACATGAAGAGTTCGCTATTG 509
194 LuLysTrpSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 210
|||||
510 AGAAACTAGTAAATAATACAGGATATACATGCGCATGGCTCGAAGAAC 559
211 LeuTyrIleuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyI 227
|||||
560 TTTATACCTACTCTCGTGGAGGTGGGACGATGATGATGATTTGGCGCAT 609
227 eLysTyrSer.Phe..... 231
|||||
610 CAATATATCTTTACATTTGAACTTCAGATGAGGCGACATACGAGATTCT 659
232 .....ThrSerAsnProProValGluLysLeuLeuProLe 243
|||||
660 TGGTGGCGGAGCGTTTACATCAAAACCCAGCTGTAGAGAAAGCTTTGGCGCT 709
243 uSerLeuLys 246
|||||
710 GTCTCTTAAAA 719
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AA135760
seq_documentation_block:
ID AA135760 standard; cDNA; 927 BP.
AC AA135760;
XX
XX 09-NOV-1996 (first entry)
XX
XX Rat mature carboxypeptidase B cDNA.
XX
XX Carboxypeptidase-B; pro-enzyme; protease; insulin; ss.
XX
XX Rattus sp.
XX
XX Key Location/Qualifiers
XX FT primer_bind complement (1..50)
XX FT /tag= a
XX FT /note= "Mature CPB 5'-end primer"
XX FT primer_bind 898..927
XX FT /tag= b
XX FT /note= "CPB 3' end primer"
XX
XX WO9623064-A1.
XX
XX 01-AUG-1996.

```

```

PF 25-JAN-1996; 96MO-US00995.
XX
XX 25-JAN-1995; 95US-0378233.
XX
XX (BIOR-) BIOTECHNOLOGY GEN CORP.
XX
XX Fulga N, Gorecki M, Hartman J, Mendelovitch S;
XX
XX WPI; 1996-362688/36.
XX
XX P-PSDB; AAW00602.
XX
XX Purified active recombinant carboxypeptidase B prodn. - by
XX
XX expressing DNA encoding the pro-enzyme, folding and enzymatic
XX
XX cleavage to give active form, used e.g. for insulin prodn.
XX
XX Example 1; Page 37-38; 49pp; English.
XX
XX A cDNA sequence (AA135760) codes for the mature form (AAW00602) of
XX
XX rat carboxypeptidase B (CPB). A full-length sequence including
XX
XX cDNA (AA135759) coding for the activation peptide (AAW00601) can
XX
XX be obt'd. by PCR amplification (see also AA135756 and AA135758) of cDNA
XX
XX derived from Sprague-Dawley rat pancreas. Pro-CPB can be expressed
XX
XX in host cells, e.g. Escherichia coli, and subsequently recovered,
XX
XX refolded and cleaved with trypsin to yield the active enzyme. CPB
XX
XX produced this way is cheaper than porcine pancreatic enzyme, and is
XX
XX free of other proteases.
XX
XX Sequence 927 BP; 253 A; 226 C; 213 G; 235 T; 0 other;

```

```

alignment_scores:
Quality: 610.00 Length: 269
Ratio: 3.389 Gaps: 5
Percent Similarity: 66.914 Percent Identity: 44.981

```

```

alignment_block:
US-09-980-881-4 x AA135760 ..

```

```

Align seg 1/1 to: AA135760 from: 1 to: 927

```

```

1 AAlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
|||||
1 GCAAGTGGACACAGCTACACCAAGTACCAACAGTGGGAACGATTGAGGC 50
17 rTyrIleGluPheIleThrGluArgHisProAspMetLeuThrIleH 34
|||||
51 GTGATTTCAACAAGTGGCACTGATTAATCCAGACTTCTCTCAGACGC 100
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||
101 TCATTTGGAACCAATTTGAAGAGGTAAACATGTAATGTCCTCAAGATT... 147
51 GlyLysGluGlnThrAlaLysAsnAlaIleThrPheLysPyrGlyIleH 67
|||||
148 GGTAAACCTAGACCGAATAGCGTGCATCTTCATTCGATTTGGTTTCA 197
67 sAlaArgGluTyrPheSerProAlaPheCysLeuThrPheIle..... 81
|||||
198 TGCAGAGAGTGCATTTCTCTGCATTTCTGCAGTGGTTTGTGAGAGAGG 247
81 .....
248 CTGTCCGTAAGCTTAATATCAAGATCCACATGAAGAACGCTTCTAGATGA 297
82 .....GlyHis..... 83
298 CTGATTTCTATGTTCTGCTGTGCTCAACATGTGATGCTTGTCTACAC 347
84 .....AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnA 97
348 CTGACATAGACAGAAATGTGTGAGAAAAACCCGCTCTACTGTGCGTGA 397
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrp 113

```



```

152 lelysalatyrillesermethisertyiserglnhisilevalphepro 168
      |||||||.....:|||||.....:|||||
560 TCAAGCATACGTGACGATCATCATCATCATCATCATCATCATCATCAT 609
      |||||||.....:|||||.....:|||||
169 tyrsertyrthrargserlyserlyasphislucluuserleuva 185
      |||||||.....:|||||.....:|||||
610 TATTCCTATGATTACAAATCTCCCGAGAACATCTGATTGAATTAACCT 659
      |||||||.....:|||||.....:|||||
185 lalaserglualaValargalaileglulysThrserlyasanthrargt 202
      |||:|||||.....:|||||.....:|||||
660 GCCTAAGCTGCCCTGAAAGAACTT...GCTACACTGATGGCACCAGT 706
      |||:|||||.....:|||||.....:|||||
202 ytrhrhisglYhisglYserglulwthrleuYtrleuAlaproglYglY 218
      |||||||.....:|||||.....:|||||
707 ACAATACAGGCGCCAGAGTACAAACATCATCTCTGCTGGGGGCTCT 756
      |||||||.....:|||||.....:|||||
219 AspAspTrpIleYrAspLeuGlYIleYsTySerPheThr 232
      |||||||.....:|||||.....:|||||
757 GATGACTGGGCTTATWAGCAAGAAATCAATATTCCTTACC 798
      |||||||.....:|||||.....:|||||

```

seq_name: /sids1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:AAQ90600

seq_documentation_block:

ID AAQ90600 standard; DNA; 1215 BP.

```

XX AAQ90600;
XX
XX 13-MAR-1996 (first entry)
XX
XX Porcine Tyr-His-Met procarboxypeptidase B coding sequence.
XX
XX Procarboxypeptidase B; carboxypeptidase B; Pichia; PCPB;
XX human serum albumin; premating factor alpha; mating factor alpha;
XX PROCBB; ds.
XX
XX Sus scrofa.
XX
XX
XX Key Location/Qualifiers
XX CDS 1..1215
XX /product= Porcine procarboxypeptidase B.
XX
XX W09514096-A1.
XX
XX 26-MAY-1995.
XX
XX 16-NOV-1994; 94MO-US13142.
XX
XX 16-NOV-1993; 93US-0153258.
XX
XX (ELIL ) LILLY & CO ELI.
XX
XX Fayerman JT, Greenen DP, Hersberger CL, Larson JL;
XX Sterner JL, Zhang H;
XX
XX WPI: 1995-200386/26.
XX P-PSDB: AAR75131.
XX
XX DNA encoding porcine carboxypeptidase B - used for transforming
XX host cells, partic. Pichia species, for prodn. of the enzyme
XX
XX Claim 2: Page 20-21; 34pp; English.
XX
XX The porcine carboxypeptidase B coding sequence can be place in a
XX bacterial or pref. Pichia yeast expression vector. The expression
XX vector further comprises the signal peptide of either human serum
XX albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
XX (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
XX pFJ474 - NRRL B-21032) or the porcine PROCBB signal peptide,
XX (designated pLGD27 - NRRL B-21027). The method can be used for
XX producing large amounts of porcine carboxypeptidase B and when
XX produced in Pichia yeast, the protein does not need solubilisation

```

CC or folding. The produced enzyme is then used for pref. cleaving
 CC basic residues from the carboxy terminus of proteins.
 XX
 SQ Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other;

alignment_scores:
 Quality: 582.00 Length: 264
 Ratio: 3.326 Gaps: 5
 Percent Similarity: 66.288 Percent Identity: 44.318

alignment_block:
 us-09-980-881-4 x AAQ90600 ..

Align seg 1/1 to: AAQ90600 from: 1 to: 1215

```

6 TyrGlulnYrHIsSerLeuAsnGlulYrSerTrpIleGlulPheI 22
  |||||||.....:|||||.....:|||||
310 TATGAGAACTACAAACACTGGGAAACGATCGAGCTTGGACTAAGCAAGT 359
  |||||||.....:|||||.....:|||||
22 eThrGlulArGHisProAspMetLeuThrLySileHisIleGlySerSerp 39
  |||:|||||.....:|||||.....:|||||
360 CACGAGTGAANAATCCAGACCTCATCTCTCGCACAGCCATGGGAACTACAT 409
  |||:|||||.....:|||||.....:|||||
39 heGlulYsTYrProleuYrValleuYsValSerGlyLySglulnThr 55
  |||:|||||.....:|||||.....:|||||
410 TTTTAGCAAAACAAATATATACCTCCTCAAGT...GGCAAACTGGACCA 456
  |||:|||||.....:|||||.....:|||||
56 AlalysAsnAlaIleTrpIleAspCySglYleHisAlaArgIuTrpI 72
  |||:|||||.....:|||||.....:|||||
457 AATAAGCTGCCATTTCATGAGCTGGTTTCATGCCAGAGAAATGAT 506
  |||:|||||.....:|||||.....:|||||
72 eSerProAlaPheCySleuTrpPheIle..... 81
  |||:|||||.....:|||||.....:|||||
507 TTCCCATGCAATTTCGCCAGTGTGTTGTGAGAGAGAGCTTCTCAACATGAT 556
  |||:|||||.....:|||||.....:|||||
81 ..... 81
557 GATATGAGAGTCAATGACAGAAATTCCTCAACAGCTAAGCTTTATGTC 606
  |||:|||||.....:|||||.....:|||||
82 .....GlyHis.....AsnArg 85
607 TTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGACCG 656
  |||:|||||.....:|||||.....:|||||
85 gMetTrpArgLySAsnArgSerPheTrpAlaAsnAsnHisCySileGlyT 102
  |||:|||||.....:|||||.....:|||||
657 AATGTGGAGAAAGAACCCGCTCTACCAATGCTGGAACACTGCATGTGCA 706
  |||:|||||.....:|||||.....:|||||
102 hrAspLeuAsnArgAsnPheAlaSerLySHsTYrCySglulnGlyAla 118
  |||:|||||.....:|||||.....:|||||
707 CAGACCCCAACAGAAATTT...GATGCTGGGTGGTGACAACTGGAGACC 753
  |||:|||||.....:|||||.....:|||||
119 SerSerSerCySserGlulwThrTYrCySglYleuYrProGlulSerG 135
  |||:|||||.....:|||||.....:|||||
754 TCTACAGACCCCTCGATGAGACTTACGTGATCTGCTGACAGATGCA 803
  |||:|||||.....:|||||.....:|||||
135 uProGlulValLySAlaValAlaSerPheLeuArgAsnIleAsnGlnI 152
  |||:|||||.....:|||||.....:|||||
804 AAAAGAGACCAAGGCCCTGGCTGATTTATACGACAAACCTCTCTCA 853
  |||:|||||.....:|||||.....:|||||
152 lelysalatyrillesermethisertyiserglnhisilevalphepro 168
      |||||||.....:|||||.....:|||||
854 TCAAGCATACGTGACGATCATCATCATCATCATCATCATCATCATCAT 903
      |||||||.....:|||||.....:|||||
169 tyrsertyrthrargserlyserlyasphislucluuserleuva 185
      |||||||.....:|||||.....:|||||
904 TATTCCTATGATTACAAATCTCCCGAGAACATCTGATTGAATTAACCT 953
      |||||||.....:|||||.....:|||||
185 lalaserglualaValargalaileglulysThrserlyasanthrargt 202
      |||:|||||.....:|||||.....:|||||
954 GCCTAAGCTGCCCTGAAAGAACTT...GCTACACTGATGGCACCAGT 1000
      |||:|||||.....:|||||.....:|||||
202 ytrhrhisglYhisglYserglulwthrleuYtrleuAlaproglYglY 218

```


808 GCTGCTGGGGCTCTGACGACTGGGCTTATGACCAAGATCATATTC 857
 230 rphenr 232
 |||||
 858 CTTCCACC 864

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AA142497

seq_documentation_block:

ID AA142497 standard; DNA; 1053 BP.

AC AA142497;

DT 12-FEB-1997 (first entry)

DE mature HCPB-(His)6-c-myc coding sequence.

KM ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADPPT;

KM mustard-ribonuclease; antibody directed enzyme prodng therapy;

KM anti-neoplastic; prodng; reverse polarity; ion pair interaction;

KM reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;

KM Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.

OS Synthetic.

FH Key Location/Qualifiers

FT CDS 1..1053

FT /tag= a /note= "double stop codon"

FT sig_peptide 1..66

FT /tag= b /note= "PelB leader sequence"

FT mat_peptide 67..987

FT /tag= c /note= "mature HCPB"

FT misc_feature 988..1047

FT /tag= d /note= "(His)6-c-myc tag"

FT W09620011-A1.

PN 04-JUL-1996.

PD 21-DEC-1995; 95WO-GB02991.

PF 16-AUG-1995; 95GB-0016810.

PR 23-DEC-1994; 94GB-0026192.

XX (ZENE) ZENECA LTD.

PA Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;

PI Henham JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;

PI Tarragona-Fiol A, Taylorson CJ;

XX WPI: 1996-321650/32.

DR P-PSDB: AAM06173.

XX Two component system for anti-tumour therapy - comprising targeting

PT moiety linked to mutated enzyme which can transform an

PT anti-neoplastic prodng

PS Reference Example 15; Page 134-136; 182pp; English.

XX A two-component system for anti-tumour therapy comprises a targeting

CC moiety linked to a mutated enzyme which can transform an anti-neoplastic

CC prodng. The system is based on antibody directed enzyme prodng therapy

CC (ADPPT) using non-naturally occurring mutant forms of host enzymes,

CC pref. human pancreatic ribonuclease (HP-RNase), (see AA142478-83).

CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can

CC be used. The mature native HCPB coding sequence (see AA142494) was cloned

CC into the secretion vector pIC1266 (which contains a PelB translation

CC leader sequence). To help purification and detection of the expression

CC construct a C-terminal peptide tag, (His)6-c-myc was added to the enzyme.

CC The tag consists of 6 histidines, a tri-peptide linker (PPE) and a
 CC peptide sequence (EQRLSEED) from c-myc which is recognised by antibody
 CC 9E10. The C-terminal is completed by addition of an asparagine. The final
 CC vector, pIC1266, contains the present sequence which encodes the
 CC mature HCPB-(His)6-c-myc protein with PelB leader sequence.
 XX

SQ Sequence 1053 BP; 279 A; 278 C; 241 G; 255 T; 0 other;

alignment_scores:

Quality: 567.00 Length: 269

Ratio: 3.133 Gaps: 5

Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:

US-09-980-881-4 x AA142497 ..

Align seg 1/1 to: AA142497 from: 1 to: 1053

1 AlaserAlaserTyrTyrGluGluThrHisSerLeuAsnGluTyrSer 17
 |||||
 67 GCAACTGCTCACTCTTACGAGAGATACAAAGTGGGAAGATAGAGC 116
 17 rTrrPleGluPheIleThrGluArgHisProAspMetLeuThrIleH 34
 |||||
 117 TTGACTCTAACAAAGTCCGCTGAGAAATCCAGCCCTCATCTCGCAGT 166
 34 IsileGlySerSerPheGluTyrTyrProLeuTyrValLeuValSer 50
 |||||
 167 TTATCGGACCAACATTTGAGGAGCGGCTATTACTCTTAAGTT... 213
 51 GlyLysGluGlnThrAlaLysAsnAlaIleTrrPheAspCysGlyLeH 67
 |||||
 214 GGCAGAGCTGGACAAATAAGCTGCTGTTTCATGAGCTGGTTCCA 263
 67 sAlaArgLutrrPheSerProAlaPheCysLeuTrrPheIle..... 81
 |||||
 264 TGCAGAGAGTGAATTCTCTGATTCCTGCCAGTGTGTGAAGAGG 313
 81
 314 CTGTTGCTACCTATGACGTGAGATCCAAAGTACAGACTTCGACAG 363
 82GlyHis..... 83
 364 TTGACTTTTATGTCCTGCTGCTCAATATGATGCTCACTACAC 413
 84AsnArgMetTrrArgLysAsnArgSerPheTyrAlaAsn 97
 414 CTGAGCCAGAGCCGATTTGGAGAAAGATCGCTCCACCATCTGAT 463
 97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrr 113
 ::|||
 464 CTAGCTGATTTGGCAGACAGCCCAAGAAATTTT...GATGCTGTGG 510
 114 CysGluGluGlyAlaSerSerSerCysSerGlnTrrTyrCysGlyLe 130
 |||||
 511 TGTGAATAATTTGGAGCTTCGAAACCCCTGATGAAATTAATCTG 560
 130 uTrrProGluSerGluPrrGluValLysAlaValAlaSerPheLeuArg 147
 |||||
 561 TCCGAGAGTCTGAAAGAGAGCCAGCCCTGCTGATTTCAATCCGA 610
 147 rGAsnIleAsnGlnIleLysAlaTrrLysSerMetHisSerTyrSerGln 163
 ::|||
 611 ACAAGTCTCTTCATCAAGGATATCGAATCCATCTGATCCCA 660
 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAsnPHsG1 180
 ::|||
 661 ATGATGATCTACCTTACTATGCTTCAAAACCTGATGATGATGAT 710
 180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysTrrs 197
 |||||


```

180 uGIuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
|||||.....:|||||.....:|||||.....:
969 TGAGTTGAATGCCCTGCTAAGCTACTGTGAAGAACTT...GCCCTAC 1015
197 eRlySAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
:::|||||.....:|||||.....:|||||.....:
1016 TGCACGGCAGCAAGTACATATGCCCCGGAGCTACACATATCTATCT 1065
214 AlaProGlyGlyGlyAspAspTyrPileTyrAspLeuGlyIleLysTyrSe 230
||| |||||.....:|||||.....:|||||.....:
1066 GCTGCTGGGGCTCTGACGACTGGCTTATGACCAAGCAATCATATTCCT 1115
230 rPheThr 232
|||||
1116 CTTCACC 1122

seq_name: /sids1/gcdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV41795
seq_documentation_block:
ID AAV41795 standard; DNA; 1263 BP.
XX AAV41795;
AC
XX
XX
XX 20-NOV-1998 (first entry)
DE Human pancreatic carboxypeptidase nucleotide sequence.
XX
XX ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KM produg therapy.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1..1248
FT CDS
FT /*tag= a
FT /product= "Pancreatic carboxypeptidase B"
FT /note= "No start codon given"
XX
XX WO9835988-A1.
XX
XX 20-AUG-1998.
XX
XX 10-FEB-1998; 98WO-GB00415.
XX
XX 29-OCT-1997; 97GB-0022727.
XX 14-FEB-1997; 97GB-0003104.
XX 18-OCT-1997; 97GB-0022003.
XX
XX (ZENE ) ZENEPA (ZENE ) ZENECA LTD.
XX
XX Edge MD;
XX
XX WPI: 1998-467168/40.
XX P-PSDB; AAW74476.
XX
XX New modified pro-domain of carboxy-peptidase B - enhances expression
XX of co-expressed proteins for production of recombinant
XX carboxy-peptidase or its fusions with antibodies, used, e.g. in
XX enzyme produg therapy
XX
XX Example 1; Page 52-53; 83pp; English.
XX
XX The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
XX modified pro-domain of CPB on a separate gene to enhance recombinant
XX expression. This process can be used to produce recombinant CPB in
XX eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
XX insulin production and protein sequencing, while its fusions with
XX antibodies are useful in antibody-directed enzyme produg therapy. The
XX modified pro-domain provides increased yields of recombinant CPB, possibly
XX by protecting the C-terminus against enzymatic degradation or by
XX increasing intracellular trafficking.
XX
XX Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;

```

```

alignment_scores:
  Quality: 567.00      Length: 269
  Ratio: 3.133
  Percent Similarity: 67.286      Percent Identity: 41.264

alignment_block:
US-09-980-881-4 x AAV41795 ..
Align seg 1/1 to: AAV41795 from: 1 to: 1263

1 AlaSerAlaSerTyrTyrGluGluThrHisSerLeuAsnGluLysTyrSe 17
|||||.....:|||||.....:|||||.....:
325 GCAACAGCAGCAGCTTATGAGAGATCAACAAGTGGGAACAGATGAGCG 374
17 rTyrPileGluPheIleThrGluArgHisProAspMetLeuThrIleH 34
||| |||.....:|||||.....:|||||.....:
375 TTGAGCTCAACAGTCCGCCCTGAGAAATCCAGCCCTCATCTCGCAGTG 424
34 lAlleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||.....:|||||.....:|||||.....:
425 TTATCGAACCACATTTGAGGAGCGGCTATTTACCTCTGAAGTT... 471
51 GlyLysGluGluThrAlaLysAsnAlaIleTyrPileAspCysGlyIleH 67
||||| ||| |||||.....:|||||.....:|||||.....:
472 GCAAAAGCTGCAAAATAATAGCTGCCATTTTCATGAGCTGGTTTCCA 521
67 sAlaArgLutThrPileSerProAlaPheCysLeuThrPheIle..... 81
|||||.....:|||||.....:|||||.....:
522 TGCCAGAGATGATTTCTCGCATTCGCCAGTGTTGTAGAGAGAG 571
81 .....:|||||.....:|||||.....:
572 CTGTTCGACCTATGAGAGCTGAGATCCAAGTGCAGAGCTTCTGCAAG 621
82 .....:|||||.....:|||||.....:
622 TTAGCTTTTATGTCCTGCTGCTGCTCAATATGATGCTCATCTACAC 671
84 .....:|||||.....:|||||.....:
672 CTGACCAAGAGCGCATTTTGGAGAAAGCTGCTCCACCCATCTGAT 721
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTyr 113
::: |||||.....:|||||.....:|||||.....:
722 CTAGCTGATTTGGCAGACAGCCCAAGAAATTTT...GATGCTGTTGG 768
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
769 TGTGAATTTGGAGCCTCTCGAAACCCCTGTGATGAACCTTACTGTGACC 818
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
819 TGCCGACAGTCTGAAAGAAACCAAGCCCTGCTATTTCATTCGCA 868
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
|||||.....:|||||.....:|||||.....:|||||.....:
869 ACAAACTCTCTTCATCAAGCATATCTGACATGCAATCGATCCCA 918
164 HisIleValPheProTyrSerTyrThrArgSerLysLysAspHisGlu 180
|||||.....:|||||.....:|||||.....:
919 ATGATGATCTACCCCTTACTCATATGCTTACAACACCGTGACAAAC 968
180 uGluSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
|||||.....:|||||.....:|||||.....:
969 TGAGTTGAATGCCCTGCTAAGCTACTGTGAAGAACTT...GCCCTAC 1015
197 eRlySAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
|||||.....:|||||.....:|||||.....:
1016 TGCACGGCAGCAAGTACATATGCCCCGGAGCTACACATATCTATCT 1065
214 AlaProGlyGlyGlyAspAspTyrPileTyrAspLeuGlyIleLysTyrSe 230

```

```

|||||
1066 GCTGCTGGGGCTCTGACGACTGGCTTATGACCAAGAAATCAAGATATTC 1115
230 rphenr 232
|||||
1116 CTCACCC 1122

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:AA142506

seq_documentation_block:

ID AA142506 standard; DNA: 1284 BP.

AA142506;

17-FEB-1997 (first entry)

PROHCPB gene with PeLB leader sequence.

ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;

mustard-ribonucleotide; antibody directed enzyme prodng therapy;

anti-neoplastic; prodng; reverse polarity; ion pair interaction;

reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;

Fd; F(ab')₂; PeLB; leader; human carboxypeptidase B; ss.

Synthetic.

Location/Qualifiers

1..1275

/*tag- a

sig_peptide

1..66

/*tag- b

sig_peptide

67..351

/*tag- c

mat_peptide

352..1272

/*tag- d

/note- "mature HCPB sequence"

MO9620011-A1.

04-JUL-1996.

21-DEC-1995; 95WO-GB02991.

16-AUG-1995; 95GB-0016810.

23-DEC-1994; 94GB-0026192.

(ZENE) ZENECA LTD.

Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;

Henham JF, Hennequin LFA, Marsham PR, Radin BR, Slater AM;

Tarragon-Floel A, Taylorson CJ;

WPI; 1996-321650/32.

P-PSDB; AAW06175.

Two component system for anti-tumour therapy - comprising targeting

modely linked to mutated enzyme which can transform an

anti-neoplastic prodng

Reference Example 18; Page 140-142; 182pp; English.

A two-component system for anti-tumour therapy comprises a targeting

modely linked to a mutated enzyme which can transform an anti-neoplastic

prodng. The system is based on antibody directed enzyme prodng therapy

(ADEPT) using non-naturally occurring mutant forms of host enzymes,

CC pref. human pancreatic ribonuclease (HP-RNase), (see AA142478-83).

CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can

be used. The present sequence is a cloned pro-HCPB gene contained in

plasmid pIC1178 and which can be expressed in E. coli.

Sequence 1284 BP; 355 A; 315 C; 299 G; 315 T; 0 other;

alignment_scores:

Quality: 567.00 Length: 269
Ratio: 3.133 Gaps: 5
Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:

US-09-980-881-4 x AA142506 ..

Align seg 1/1 to: AA142506 from: 1 to: 1284

```

1 AlaSerAlaSerTyrTyrGluGluTyrHisSerLeuAsnGluLeuTyrSe 17
|||||
352 GCACAGGACACAGTATGAGAGTACAAAGAGTGAAGACATACAGGC 401
17 rTTPleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
|||||
402 TTGGAGCTCAACAGATGCGCCAGATCCAGCCCATCTCTCGCAGTCG 451
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||
452 TTAGCGAACACATTTGAGGAGCGGCTATTACTCTGAAAGTT... 498
51 GlyLysGluGluThrAlaLysAsnAlaIleTyrPleAspCysGlyLeuH 67
|||||
499 GGCAGAGCTGACAAATATAGCCTGCCATTTCATGAGCTGAGTTTCA 548
67 sAlaArgGluTyrPleSerProAlaPheCysLeuTyrPheIle..... 81
549 TGCCAGAGAGTGAATTCCTGCAATTCGCCAGCGTTTGAAGAGAGC 598
81 ..... 81
599 CTTGTCGACGATGAGCTGAGATCCAGTACAGAGCTTTCGACAG 648
82 .....GlyHis..... 83
649 TTAGACTTTATGCTCGCTGCTGCATATTTGATGGCTATACAC 698
84 .....AsnArgMetTyrPheLysAsnArgSerPheTyrAlaAsnA 97
699 CTGACCAAGAGCCGATTTGGAGAAAGCTGCTCCACCCATCTGAT 748
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTyr 113
:: |||||
749 CTAGCTGATTTGCAACAGACCCCAAGAAATTT...GATCTGTTGG 795
114 CysGluGluGlyAlaSerSerSerSerCysSerGluTyrCysGlyLe 130
|||||
796 TGTGAATTTGGAGCTCTCGAAACCCCTGTGATGAACCTTACTGTGACC 845
130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
846 TGCCCCAGAGTCTGAAAGAGAGACCAAGCCCTGGCTAATTCATCA 895
147 rGAsnIleAsnGluIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
|||||
896 ACAAACTCTCTTCATCAAGGATATCGACAAATCTCTGATCTGACCCA 945
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 180
|||||
946 ATGATGATCTACCCCTTACTCATATGCTTCAAAACCGGTGAGAACATGC 995
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
|||||
996 TGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAGAATCT...GCCCTAC 1042
197 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
|||||
1043 TGACAGGACCAAGATACATATGCGCGGAGCTACCAACATATTCCT 1092
214 AlaProGlyGlyGlyAspAspTyrPleTyrAspLeuGlyIleLysTyrSe 230

```

|||||
1093 GCTGCTGGGGCTCTGACGACTGGCTTATGACCAAGATCATATTTTC 1142
230 rphetr 232
|||||
1143 CTCACCC 1149

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAx24804

seq_documentation_block:

ID AAx24804 standard; DNA; 1053 BP.

AC AAx24804;

XX 21-JUN-1999 (first entry)

XX Human carboxypeptidase B (D253K mutant)-6His cmc DNA.

XX Carboxypeptidase B; mutant; human; T7 promoter; operator;

XX gene expression; vector; pZK7#3.3; ds.

XX Homo sapiens.

XX MO9905297-A1.

XX 04-FEB-1999.

XX 21-JUL-1998; 98WO-GB02175.

XX 25-JUL-1997; 97GB-0015660.

XX (ZENEC) ZENEC LTD.

XX Bundell KR, Hockney RC, Kara BV, Ploil D;

XX WPI; 1999-142947/12.

PT New system for expression of recombinant proteins - comprises a T7
PT based promoter-driven expression system having upstream and
PT downstream operator sequences

P5 Disclosure; Page 30; 56pp; English.

CC This is the coding sequence of a mutant human carboxypeptidase
CC (D253K) with C-terminal hexahistidine cmc tag. The sequence has
CC been placed downstream of the Erwinia carotovora pelb secretory
CC leader sequence and cloned into pET11a and into novel, claimed
CC vector pZT7#3.3. pZT7#3.3 has a novel T7 based promoter-driven
CC expression system that utilizes upstream and downstream perfect
CC palindromic operator (pPO) sequences. Accumulation of biologically
CC active CPB(D253K)-6His-cmc in E. coli transformants could be
CC raised to higher levels using the pZT7#3.3 system compared with
CC the pET11a system using appropriate levels of IPTG inducer. The
CC new T7 promoter-based expression system provides improved control
CC of expression and improved levels of protein expression. Basal
CC expression in the absence of inducer can be reduced to a level
CC which permits the cloning and expression of toxic gene products.
CC The system also allows control of production of heterologous
CC proteins in an inducer concentration-dependent manner over a wide
CC range of expression levels.

XX Sequence 1053 BP; 281 A; 277 C; 240 G; 255 T; 0 other;

alignment_scores:

Quality: 560.00 Length: 269
Ratio: 3.111 Gaps: 5
Percent Similarity: 66.914 Percent Identity: 40.892

alignment_block:

US-09-980-881-4 x AAx24804 ..

Align seg 1/1 to: AAx24804 from: 1 to: 1053

1 AAlaserlaserTyrTyrGlutTyrHisSerLeuasnGluTyrSe 17
|||||
67 GCAAGCTGCTCCTCTTACGAGAACTACAACTGCAAGCAATAGAGGC 116
17 rTPrILeGluPheLeThrgLuarGHisProAspMetLeuThrLysLeh 34
|||||
117 TTGAGCTCAACAACTGCGCCTGAGAACTCCAGCCCTCTCTCGCAGTG 166
34 IsileGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||
167 TTATCGGAGACCAATTTGAGGAGCGGCTATTTACCTCTTAAGTT... 213
51 GlyLysGluGlnThrAlaLysAsnAlaLeuTPrLeaSpCysGlyLeh 67
|||||
214 GGCAAAGCTGAGCAAAATTAAGCCCTGCAATTTTCATGAGCTGTGTTCCA 263
67 sAlaArgGluTPrLeSerProAlaPheCysLeuTPrPheLe... 81
264 TGCCAGAGAGTGATTTCTCTCTGCAATTCCTGCAAGTGTGTTAGAGAGG 313
81
314 CTGTTCGTACCTATGAGAGTGAATCCAGTGAAGAGCTTCTGACAG 363
82GlyHis..... 83
364 TTAGACTTTTATGTCCTGCTGCTGCTCAATATGATGAGCTACATCTACAC 413
84AsnArgMetTPrArgLysAsnArgSerPheTyrAlaAsn 97
414 CTGAGCAAGAGCGCATTTGGAGAAAGACCTGCCACCCACTCTGAT 463
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTPr 113
:::|||||
464 CTAGCTGATTTGGCAGACAGACCCCAAGAAATTT...GATGCTGTGG 510
114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
|||||
511 TGTGAATTTGGAGCCCTCGCAACCCCTGTGATGAATCTACTGTGAGC 560
130 uTyrProGluSerLupProGluValLysAlaValAlaSerPheLeuArg 147
|||||
561 TGCCGAGAGTGTGAAAAGAGACCAAGCCCTGCTGATTTCAATCCGCA 610
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
:::|||||
611 ACAAACTCTCTCCATCAAGCATATCTGACAAATCCACTGACTCCCA 660
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG1 180
|||||
661 ATGATGATCTACCTTACTGATGCTTACCAACTCGGTGAGCAATATGC 710
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
|||||
711 TGAATTAATGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 757
197 eTyrAsnThrArgTyrThrHisGlyHisSerGluThrLeuTyrLeu 213
|||||
758 TGCAAGGACCAATATACATATGAGCCGAGAGCTACCAACATATATCT 807
214 AlaProGluGlyLysAspTPrIleTyrAspLeuGlyLysTyrSe 230
|||||
808 GCTGCTGGGGCTCTAAAGACTGGCTTATGACCAAGCAATACAGATATTC 857
230 rPheThr 232
858 CTCACCC 864

This Page Blank (uspto)

OM of: US-09-980-881-4 to: Issued_Patents_NA.* out_format : pfs

Date: Sep 18, 2002 6:56 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=x1h
-O/cgn2_1/USPTO.spool/US09980881/runat.16092002.140129.10653/app-query.fasta_1.1126
-DB=Issued_Patents_NA -OPMT=fastap -SUFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MIMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 -CGN1_1.62 -NCPU=6 -ICPU=3 -LONGLOS
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-4
Query length: 246
Database: Issued_Patents_NA.*
Database sequences: 383533
Database length: 122816752
Search time (sec): 90.520000

score_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation	..	1.8e-147	1272	1
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-869-057-1 +		1272.00	2785.71	1.8e-147	1272	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-649-591B-2 +		1272.00	2781.78	2.9e-147	1749	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-277-540-2 +		1272.00	2781.78	2.9e-147	1749	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-430-787A-2 +		1272.00	2781.78	2.9e-147	1749	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-782-760-5 +		610.00	1321.16	6.7e-66	927	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-696-139-3 +		582.00	1259.13	1.9e-62	921	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-696-139-1 +		582.00	1255.71	3.0e-62	1215	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-67 +		567.00	1224.86	1.5e-60	999	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-64 +		567.00	1224.21	1.7e-60	1053	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-56 +		567.00	1221.76	2.3e-60	1284	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-71 +		567.00	1221.76	2.3e-60	1059	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-74 +		559.00	1206.39	1.7e-59	1059	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-860-882A-77 +		559.00	1206.39	1.7e-59	1059	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-171-945-112 +		549.00	1177.19	8.8e-58	1870	1			
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-171-945-124 +		549.00	1175.44	8.8e-58	2154	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-640-906-1 +		465.50	996.87	7.7e-48	1257	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-395-936-1 +		465.50	996.87	7.7e-48	1257	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-640-906-3 +		459.00	982.51	4.9e-47	1251	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-395-936-3 +		459.00	982.51	4.9e-47	1251	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-998-416-125 -		169.50	351.30	7.0e-12	515	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-331-709-2 +		163.00	335.18	5.6e-11	591	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-706-216-1 +		96.50	168.83	0.1026	2719	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-484-105-15 +		95.00	166.52	0.1379	2504	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-484-106-15 +		95.00	166.52	0.1379	2504	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-689-974-2 -		88.50	157.82	0.4211	1576	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-058-376-2 -		88.50	157.82	0.4211	1576	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-233-989-1 +		88.50	152.42	0.8408	2439	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-435-262-1 +		88.50	152.40	0.8429	2443	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-734-550-1 +		88.50	152.40	0.8429	2443	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-073-297-1 +		88.50	152.40	0.8429	2443	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-609-049A-27 +		86.50	135.27	7.59	6831	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-170-996-27 +		86.50	135.27	7.59	6831	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-966-363-10 -		85.00	144.26	2.40	2520	1			
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-969-683A-10 -		85.00	144.26	2.40	2520	1			
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-297-928-6 -		85.00	144.26	2.40	2520	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-903-910-1 +		82.50	157.80	0.4219	537	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-255-966B-1 +		82.50	157.80	0.4219	537	1			
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-222-638B-1 +		82.50	157.80	0.4219	537	1			

/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-920-812-13 +	81.50	120.09	53.18	9515
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-920-827-13 +	81.50	120.09	53.18	9515
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-921-177-13 +	81.50	120.09	53.18	9515
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-362-577C-13 +	81.50	120.09	53.18	9515
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-920-828-13 +	81.50	120.09	53.18	9515

seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-869-057-1

seq_documentation_block:

Sequence 1, Application US/08869057
Patent No. 5985562
GENERAL INFORMATION:
APPLICANT: Morser, Michael J
TITLE OF INVENTION: Method of Detecting Thrombotic Disease
TITLE OF INVENTION: Risk
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Berlex Biosciences Legal Department
STREET: 15049 San Pablo Avenue
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94604-0099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,057
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Washien, Wendy L.
REGISTRATION NUMBER: 36,301
REFERENCE/DOCKET NUMBER: 51509AUSM1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-262-5411
TELEFAX: 510-262-7095
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1272 base pairs
TYPE: nucleic acid
TOPOLOGY: double
MOLECULE TYPE: linear
PUBLICATION INFORMATION:
AUTHORS: Eaton, Dan L.
AUTHORS: Malloy, Beth E.
AUTHORS: Tsai, Siao P
AUTHORS: Henzel, William
AUTHORS: Dryden, Dennis
TITLE: Isolation, Molecular Cloning, and Partial
TITLE: Characterization of a No. 5985562el Carboxypeptidase B
JOURNAL: J. Biol. Chem.
VOLUME: 266
ISSUE: 32
PAGES: 21833-21838
DATE: No. 5985562 15-1991
US-08-869-057-1
alignment_scores:
Quality: 1272.00 Length: 301
Ratio: 5.171 Gaps: 3
Percent Similarity: 81.728 Percent Identity: 81.395
alignment_block:
US-09-980-881-4 x US-08-869-057-1 ..
Align seq 1/1 to: US-08-869-057-1 from: 1 to: 1272

```

246 s 246
1
1243 A 1243

seq_name: /cgn2_6/ptcddata/1/ina/5A_COMB.seq:US-07-649-591B-2

seq_documentation_block:
; Sequence 2, Application US/07649591B
; Patent No. 5206161
;
GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
ATTORNEY/AGENT INFORMATION:
; NAME: Haasek, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9681
; TELEX: 910/371-7168
;
INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
;
FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
;
FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 41 to 106
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
;
US-07-649-591B-2

alignment_scores:
Quality: 1272.00 Length: 301
Ratio: 5.171 Gaps: 3
Percent Similarity: 81.728 Percent Identity: 81.395

alignment_block:
US-07-649-591B-2 x US-07-649-591B-2 ..

Align seg 1/1 to: US-07-649-591B-2 from: 1 to: 1749

```

```

1 AlaserAlaserTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
  |||||||
383 GCCTCCGCAATCGTACTATGACAGTATCACTCAATAATGAATCTATATC 432
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
  |||||||
433 TTGGATGAAATTTTATTAACAGAGGACATCTGTATGCTTACAAAAATCC 482
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
  |||||||
483 ACATTGGATTCCTCATTTGAGAACTACCACTCATGTTTAAAGGTTTCT 532
51 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
  |||||||
533 GGAAGAGAACAAACAGCAAAATGCCATGATGCTGATGCAATGCCA 582
67 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 83
  |||||||
583 TGCAGAGAAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 632
83 .....
633 TAACCAATTCTATGGGATATAAGGCAATATACCAATCTCTGAGGCTT 682
83 .....
683 GTCGATTTCTATGTATGCCGGTGTAAATGTGACGGTTATGACTACTC 732
84 .....
733 ATGGAAGAAAGAAATCGAATGTGGAGAAAGAAACGTTCTTCTATGCGACA 782
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrp 113
  |||||||
783 ATCATGTGCAATCGSACAGACACTGAATAGAACTTGTCTTCCAAACACTGG 832
114 CysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLe 130
  |||||||
833 TGTGAGAGAGGTGCATCCAGTTCCTCATCTCGGAAACACTACTGTGACT 882
130 uTyrProGluSerGluProGluValAlaValAlaAlaSerPheLeuArg 147
  |||||||
883 TTATCTGATCGAAGACAGAACTGAAGGAGAGTGTCTTCTTGAGAA 932
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
  |||||||
933 GAATATTCACACGATTAAGCATATCATGATGATTCATATCTCCAG 982
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 180
  |||||||
983 CATATAGTGTTCATTCATTCCTATACAGCAATGAAAAAGCAAGACATGA 1032
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
  |||||||
1033 GGAAGCTGTCTAGTATGCCAGTGAAGCAGTTCGTCTATTGAGAAAACTA 1082
197 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
  |||||||
1083 GTAAATAATCCAGTATACATGCGCATGGCTCAGAAACCTTATACCTA 1132
214 AlaProGluGlyLysAspTrpIleTyrAspLeuGlyLysTyrSe 230
  |||||||
1133 GCCTCGAGAGGTGGGACGATGTGATCTATGATTTGGCATCAAAATATTC 1182
230 r.Phe..... 231
1183 GTTTCATTAATGAATCTCGAGATACGGGCAATACGATTTCTGCTGCCGG 1232
232 .....ThisSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
  |||||||
1233 AGGGTATACATCAACCAACCTCTGTAGAGAACCTTTTGGCGGTCTGTCTAAA 1282

```

```

246 s 246
1283 A 1283
seq_name: /cgn2_6/pdata/1/lna/5A_COMB.seq:US-08-277-540-2
seq_documentation_block:
; Sequence 2, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: NO. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: palin (genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-277-540-2

alignment_scores:
Quality: 1272.00 Length: 301
Ratio: 5.171 Gaps: 3
Percent Similarity: 81.728 Percent Identity: 81.395

alignment_block:
US-09-980-881-4 x US-08-277-540-2 ..
Align seg 1/1 to: US-08-277-540-2 from: 1 to: 1749

1 AlaserAlaserTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
  |||||||
383 GCCTCCGCAATCGTACTATGACAGTATCACTCAATAATGAATCTATATC 432
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
  |||||||
433 TTGGATGAAATTTTATTAACAGAGGACATCTGTATGCTTACAAAAATCC 482

```

```

34 1st1leGlySerSerPheGluLysTyTProLeuTyValLeuLysValSer 50
|||||
483 ACATTTGGATCCCTCATTTTGAGAACTACCCACTATGTTTAAAGCTTTC 532
51 G1yLysG1uGlnThrAlaLysAsnAla1leTrrP1leAspCySGly1leH1 67
|||||
533 GGAAGAAAGACAAACAGCCAAAATGCCATATGGATTGACGTGGAAATCCA 582
67 sAlaArgGluTrrP1leSerProAlaPheCySLeuTrrPhe1leG1yH1s. 83
|||||
583 TGCCAGAGAAATGATCTCTCCCTCTTCTGCTGTGTTCAATAGGCCATA 632
83 ..... 83
633 TAACCAATTCTATGGGATATAAGGCCAATATACCAATCTCTGAGGCTT 682
83 ..... 83
683 CTGCATTTCTATGTTATGCCGGGGTTAATGTGGACGGTTATGACTACTTC 732
84 ..... 84
733 ATGGAAGAAAGAAATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGACA 782
97 sNH1sCyS1leG1yH1rAspLeuAsnArgAsnPheAlaSerLysH1sTrr 113
|||||
783 ATCATATTCGACGAAACAGACTGGAATAGGAAGACTTGTCTTCCAAACACTGG 832
114 CySGluGluG1yAlaSerSerSerSerCySergLutHrTyCySG1yLe 130
|||||
833 TCGTAGGAAGGTGCATCCAGTTCCATCTCTCGAAACCTACTGTGCACT 882
130 uTyTrrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
|||||
883 TTATCTCTGAGTCAGAACCGAAGTAGGAGCACTGCTATGTTCTTGAGAA 932
147 rGAsn1leAsnGln1leLysAlaTyTrr1leSerMetH1sSerTySergln 163
|||||
933 GAAATATTCACCAAGATTAAAGCATATCATGATCATTCATCTCCAG 982
164 H1s1leValPheProTySerTyTrrH1rAspSerLysSerLysAsnH1sG1 180
|||||
983 CATATAGTGTTCATATTCCTATACAGAACTAAAGAAAGACCATGCA 1032
180 uGluLeuSerLeuValAlaSerGluAlaValArgAla1leGluTyS1Hs 197
|||||
1033 GGAAGTCTCTAGTAGCCAGAGCAAGTCTGCTATTTGAGAAACTA 1082
197 eTlyAsnThrArgTyTrrH1sG1yH1sG1ySerGluThrLeuTyLeu 213
|||||
1083 GTAAATAATACCAAGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 1132
214 AlaProGlyGlyGlyAspAspTrrP1leTyTrrAspLeuGly1leTyS1e 230
|||||
1133 GCTCCTGGAGGTGGGAGCATTTGATGATTTGGGCATCAATAATATTC 1182
230 r.Phe..... 231
1183 GTTACCAATTGAACCTCGAGATACGGGACATACGATTCCTGCTGCCGG 1232
232 .....ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuTy 246
|||||
1233 AGCGTTACATCAAAACCCAGCTGTAGAGAAAGCTTTTGGCGGTGTCTTAA 1282
246 s 246
1283 A 1283
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-430-787A-2
seq_documentation_block:
; Sequence 2, Application US/08430787A
; Patent No. 5593674

```

```

; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Raton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Gentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 Inch, 360 KB floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Gentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-430-787A-2

alignment_scores:
    Quality: 1272.00      Length: 301
    Ratio: 5.171          Gaps: 3
    Percent Similarity: 81.728    Percent Identity: 81.395

alignment_block:
US-09-980-881-4 x US-08-430-787A-2 ..

Align seg 1/1 to: US-08-430-787A-2 from: 1 to: 1749

1 1AlaSerAlaSerTyTrrGluGlnTyH1sSerLeuAsnGln1leTyS1e 17
|||||
383 GCCTCCGCATCTGCTATGAGACAGTATCAGCTAATGAATGAATCTATTC 432
17 tTrrP1leGluPhe1leThrGluArgH1sProAspMetLeuThrLys1leH 34
|||||
433 TTGATAGAAATTATTAAGTGAAGCATCCGATATGCTTACAAAATTC 482
34 1st1leGlySerSerPheGluLysTyTProLeuTyValLeuLysValSer 50
|||||
483 ACATTTGGATCCCTCATTTTGAGAAAGTACCCACTATGTTTAAAGCTTTC 532
51 G1yLysGluGlnThrAlaLysAsnAla1leTrrP1leAspCySGly1leH1 67
|||||
533 GGAAGAAAGACAAACAGCCAAAATGCCATATGGATTGACGTGGAAATCCA 582

```

```

      NUMBER OF SEQUENCES:   8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036
```

COMPUTER READABLE FORM:

```
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/M5-POS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/782,760  
FILING DATE: 13-JAN-1997  
CLASSIFICATION: A35  
  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/378,233  
FILING DATE: 25-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0336/43847  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO.: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 927 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOCHETICAL: NO  
ANTI-SENSE: NO  
  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..927  
US-08-782-760-5
```

alignment_scores:

```
Quality:    610.00          Length:       269  
Ratio:     3.389           Gaps:         5  
Percent Similarity: 66.914 Percent Identity: 44.981
```

alignment_block:

```
US-09-980-881-4 x US-08-782-760-5 ..  
  
Align seg 1/1 to: US-08-782-760-5 from: 1 to: 927
```

```
1 AlaseralaserTYTGTGUGLINTYRHISserLeuansgILUetyrse 17  
|||||...|| :||:::||::: || |::  
1 GCAGATGGACACACTAACCAACAAGTAGCACAACTCGGAAGAAGATTGAGGC 50  
17 rTPrllequbhelletnrgIuhAghIsPrOAsPmetLeuthryslenH 34  
|||||..... ::|||:::::||||:~::~  
51 GTGGATTTCMAAGAtTGCCACTGTGATAACCAGACCTTGACTCAGAGCG 100  
34 ISlIGLSerSerpheGLIntyTyrrProLeutyValleuLyvalser 50  
|||||:::|||||:::|||||:::|||||:::  
101 TCATTGGACCAcCATTTTAAGAGAGTAAcATGtATgtCCTCAAAtt... 147  
51 glytysgluGIintThralAlalysAnaLaileTrprlleaspCySglyleHI 67  
|||||::: ~~~ |||||:::|||||:::|||||:::  
148 GSTAAAATCAGACCGAATAAGCCGCCATCTTCAATCGATTCGTTGCCA 197  
67 salaragLUirprllsesrrproLAphecylsuLeurPhelle..... 81  
138 TCCAAGAcAgTggattttctcgcatTTctgcacigtggttgtagAGAGG 247  
81 ..... 81
```

[illegible]

```

495 AGCCCCAGAGTCTGAAAAAGAGCAAGGCCCTGCAGATTTCATCCGCA 544
147 rGAsnIleAsnGlnIleIysAlaTyrIleSerMetHisSerTyrSerGln 163
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545 ACAACCTCTCCACATCAAGGCTACCTGACATCCATCATCATCACTACAG 594
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG 180
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595 ANGATGCTTACCTTACCTTACCTATGACTACAAAGCTGCGAAGCATGGA 644
180 uGluSerLeuValAlaSerGluValAlaArgAlaIleGluLysThrS 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 GGAATTGATGCCCTGCTGTAAGAGTGGCGCAAGAGACTT...GCCACTC 691
197 eLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 TGCATGGCACAAGTACATATGATGAGGAGGAGCTACAAACATCTATGCT 741
214 AlaProGlyGlyGlyAspAspTrrPleTyrAspLeuGlyIleLysTyrS 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 GCTGCTGGGGGATGTGACGACTGTCTTATGATCAGGGAATCAAAATATTC 791
230 rPheThr 232
|||||
792 CTTTACC 798

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:us-08-696-139-3

seq_documentation_block:

```

: Sequence 3, Application US/08696139
: Patent No. 5672496
: GENERAL INFORMATION:
: APPLICANT: Fayerman, Jeffrey T.
: APPLICANT: Greenen, David P.
: APPLICANT: Hersberger, Charles L.
: APPLICANT: Larson, Jeffrey L.
: APPLICANT: Sterner, Jane L.
: APPLICANT: Zhang, Haichao
: TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
: TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESS: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: United States of America
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/696,139
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/153,258
: FILING DATE: 16-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Gaylo, Paul J.
: REGISTRATION NUMBER: 36,808
: REFERENCE/DOCKET NUMBER: X-8681
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317) 276-0756
: TELEFAX: (317) 276-3861
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 921 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

```

; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..921
US-08-696-139-3

```

```

alignment_scores:
  Quality: 582.00      Length: 264
  Ratio: 3.326
  Percent Similarity: 66.288      Percent Identity: 44.318

```

alignment_block:

US-09-980-881-4 x US-08-696-139-3 ..

Align seg 1/1 to: US-08-696-139-3 from: 1 to: 921

```

6  TYRGLuInTyrHisSerLeuAsnGlnIleTyrSerTrpIleGluPheI 22
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16  TATGAGAAGTACAAACAACGTAAGAGAGGCTGAGCTGGAAGTAAAGT 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22  eThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerP 39
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66  CACCAGTGAATAATCCACAGCCTCATCTCGCACAGCCATGGAGACTCAT 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39  heGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThr 55
|||:|||||:|||||:|||||:|||||:|||||:|||||:
116  TTTTAGAAGAACAAATATATACCTCTCAAGGT...GGCAAACTGGAGACA 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56  AAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpI 72
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163  AATTAAGCTGCTGCATTTTCATGAGACTGTGGTTCATGCGAGAGATGAT 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72  eSerProAlaPheCysLeuTrpPheIle..... 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213  TTCCCATGCAATTTTGGCCAGTGTTGTGACAGAGGCTGTTCTCCTATG 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81  ..... 81
263  GATATGAGACTACATGACAGAAATTCTCAACAGCTAGACTTTATGTC 312
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82  .....GlyHis.....AsnAr 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313  TTGCGGTGCTCAATATGATGCTCATCTACACCTGAGCAAGAACCG 362
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85  gMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyT 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363  AATGTGAGAAAGACCCGCTCTACCAATGCTGGAATCACTGCAATTGGCA 412
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102  hAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAla 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
413  CAGACCCCAACAGAAATTTT...GATGCTGGGTGTCACAACTGAGGCC 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119  SerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerG 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460  TCTACAGACCCCTGCGATGAGACTTACTGTGGACTCTCGACGATGTGGA 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135  uProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnI 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510  AAAAGAGACCAAGCCCTGCTGCTGATTTATACGCAACACTCTCTCCCA 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152  leLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560  TCAAGCATACCTGACATCAGTACATACACAGATGATCTTACCT 609
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169  TyrSerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuVal 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610  TATTCCTATGATTAACAACCTCCCGAGAACATCTGATGGAATTAACCT 659
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185  lAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660  GCCTAAGCTGCGCTGGAAGAACTT...GCTACACTGTATGCGACCACT 706
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 1053 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-64
```

```
alignment_scores:
Quality: 567.00 Length: 269
Ratio: 3.133 Gaps: 5
Percent Similarity: 67.286 Percent Identity: 41.264
```

```
alignment_block:
US-09-980-881-4 x US-08-860-882A-64 ..
```

```
Align seg 1/1 to: US-08-860-882A-64 from: 1 to: 1053
```

```
1 AAlaserAlaserTyrTyrGluGluGlnHisSerLeuAangLylIeTyrSe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCACTGCTCACTCTTACGACAGATGACAAAGTGGAAAGCATGAGAGCC 116
17 rTrrlGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TTGACACTCAACAGTCGCCACTGAGAAATCCAGCCCTCATCTCGCAGCTG 166
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTATCGAAGACACATTTGAGGAGCGCTATTACTCTCTGAAGGT... 213
51 GLyAsGluGlnThrAlaLysAsnAlaIleTrrPheAspCysGlyIleH1 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GGCAAAGCTGACAAATAAGCCCTGCCATTTTCATGAGACTGTGTTCCA 263
67 sAlaArgLulTrrPheSerProAlaPheCysLeuTrrPheIle..... 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TGCCAGAGAGTGTCTCTCGATCTGCCAGTGTGTTGTAAGAGAGG 313
81 ..... 81
314 CTGTTCGACTATGAGCGTGAATCCAGTGAAGAGCTTCTGACAGAG 363
82 .....GlyHis..... 83
364 TTAGACTTTTATGTCTGCTGCTGCTCAATATGATGCTGACATCACTAC 413
84 .....AsnArgMetTrrArgLysAsnArgSerPheLysIleAsnA 97
414 CTGACCAAGAGCCGATTTTGAGAAAGACCTGCTCCACCATCATGAT 463
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrr 113
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 CTAGCTGATTTGACAGACAGCCCAAGAAATTTT...GATGCTGGTGG 510
114 CysGluGluGlyAlaSerSerSerSerSerSerSerSerSerSerSer 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511 TGTGAATTTGAGGCTCTCGAAACCCCTGTGATGAACCTTACTGTGAGCC 560
```

```
130 uTrrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 TGGCCGAGAGTGTGAAGAGAGACCAAGGCCCTGCTATTTCAATCCGCA 610
147 rGAsnIleAsnGlnIleLysAlaTrrIleSerMetHisSerTyrSerGln 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611 ACAAACTCTTCCATCAAGGCATATGTGACAAATCCACTGATCCCA 660
164 HisIleValPheProTyrSerTrrThrAspSerLysSerLysAspHisG 180
661 ATGATGATCTACCTTACTGATATGCTTACAAACTCGGAGAACAAATGC 710
180 uGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
711 TGAGTTGAATGCCCTGGCTAAAGCTAGCTGAAGAACTT...GCCCTAC 757
197 eTrrAsnThrArgTrrThrHisGlyHisGlySerGluThrLeuTrrLeu 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
758 TGCACGGGACCAAGTACACATATGCCCCGGAGCTACAAACATCTATCT 807
214 AlaProGlyGlyAlaAspAspTrrIleTrrAspLeuGlyIleLysTyrSe 230
|||:|||||:|||||:|||||:|||||:|||||:|||||:
808 GCTGCTGGGGGCTCTGACGACTGGCTTATGACCAAGAAATCAGATATTC 857
230 rPheThr 232
|||||
858 CTTCACC 864
```

```
seq_name: /cgn2_6/prodata/1/ina/5b_comb.seq:US-08-860-882A-56
```

```
seq_documentation_block:
; Sequence 56, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSER: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
```

: LENGTH: 1263 bases
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 US-08-860-882A-56

alignment_scores:
 Quality: 567.00 Length: 269
 Ratio: 3.133 Gaps: 5
 Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:
 US-09-980-881-4 x US-08-860-882A-56 ..

Align seg 1/1 to: US-08-860-882A-56 from: 1 to: 1263

```

1  AlaserAlaserTyrTrpGluGlnTyrHisLeuAsnGluLeuTyrSer 17
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 GCACACGACACAGCTTATGAGAGTACACAGTGGGAAACGATAGAGGC 374
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17  rTrpLeuGluPheLeuThrGluArgHisProAspMetLeuThrLysIleH 34
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 TTGGACTCAACAGTGGCCACAGATGACGCCCTCATCTCGCAGGTG 424
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34  IsIleGlySerSerPheGluLysTyrProLeuTyrValIleLysValSer 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 TTTATCGGACACACATTTGAGGAGCGGCTTTTACCTCCGGAAGGTT... 471
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51  GLySGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
472 GGGAAAGCTGGACAAATTAAGCTGCATTTTCATGAGCTGTTTCCA 521
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67  sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle..... 81
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
522 TGCCAGAGAGTGGATTTCCTCGCATTCGCGTGTGTGTAAGAGAG 571
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81  ..... 81
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 CTGTCGTACCTATGAGAGTGCAGTCCAGTCCAGAGCTTCGACAG 621
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82  ..... 83
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
622 TTGAGCTTTATGTCTGCTGCTGCATATTGATGGCTACATCTACAC 671
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84  ..... 97
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
672 CTGACCAAGAGCCGATTGAGAGAAAGACTCGCTCCACCATACCTGAT 721
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97  snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrp 113
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
722 CTAGCTGCATTTGGCAGACCCCAACAGAAATTTT...GATGCTGATGG 768
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 CysGluGlnGluAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
769 TGTGAAATTGGAGCTCTCGAAACCCCTGTGATGAAACTTACTGTGACC 818
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
130 uTrpProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
819 TGCCGAGAGTGTGAAAGAGACCAAGCCCTGGCTGATTTACATCCGA 868
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 rGAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
869 ACAAACTCTCTTCATCAAGCATATCTGACAATTCACCTGACTCCCAA 918
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHis 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
919 ATGATGATCTACCTTACTACTATGCTTACAAACTCGGTAGACAAATGC 968
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 uGluLeuSerLeuValAlaSerGluAlaValAlaGalaIleGluLysThr 197
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
969 TGAGTTGAATGCCCTGGCTAAGCTACTGTGAAAGAACTT...GCCAC 1015
  
```

```

197 erLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1016 TGCAGGGCACCAGTACATATGAGCCCGGAGCTGACACATCTATCT 1065
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 AlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1066 GCTCTGGGGGCTCTGACGACTGGGCTTATGACCAAGATCATATCTC 1115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
230 rPheThr 232
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1116 CTCACCC 1122
  
```

seq_name: /gen2_6/ptodata/1/ina/5B_COMB.seq:US-08-860-882A-71

seq_documentation_block:

; Sequence 71, Application US/08860882A
 ; Patent No. 5985281

GENERAL INFORMATION:

APPLICANT: TAYLORSON, CHRISTOPHER JOHN
 APPLICANT: EGELTE, HENDRIKUS JOHANNES
 APPLICANT: TARRAGONA-FIOL, ANTONIO
 APPLICANT: RABIN, BRIAN ROBERT
 APPLICANT: BOYLE, FRANCIS THOMAS
 APPLICANT: HENNAM, JOHN FREDERICK
 APPLICANT: BLAKELY, DAVID CHARLES
 APPLICANT: MARSHAM, PETER ROBERT
 APPLICANT: HEATON, DAVID WILLIAM
 APPLICANT: DAVIES, DAVID HUM
 TITLE OF INVENTION: CHEMICAL COMPOUNDS
 NUMBER OF SEQUENCES: 77
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PILLSBURY, MADISON & SUTRO
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk
 COMPUTER: IBM compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/860,882A
 FILING DATE: JUNE 23, 1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: DONALD J. BYRD

REGISTRATION NUMBER: 25,323
 REFERENCE/DOCKET NUMBER: 9901/238653
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 861-3027
 TELEFAX: (202) 822-0944

TELEEX: 6174627 CUSH
 INFORMATION FOR SEQ ID NO: 71:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1284 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-08-860-882A-71

alignment_scores:

Quality: 567.00 Length: 269
 Ratio: 3.133 Gaps: 5
 Percent Similarity: 67.286 Percent Identity: 41.264

alignment_block:

US-09-980-881-4 x US-08-860-882A-71 ..

Align seg 1/1 to: US-08-860-882A-71 from: 1 to: 1284

```

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 GCACAGGACACAGTATGAGAGATACAAACAGTGGGAACATAGAGGC 401
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrIleHis 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 TTGGGACTCAACAAGTCGCCACTGAGAAATCCACCCCTCATCTCCGAGTG 451
34 IsileGlySerSerPheGluTyrTyrProLeuTyrValLeuIysValSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 TTATCGAACCACATTTGAGGAGCGGCTATTACTCTCGAAGGT... 498
51 GlyIysGluGlnThrAlaIysAsnAlaIleTrpIleAspCysGlyIleH 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
499 GGCAAAGCTGACAAATTAAGCCCTCATTTTCATGAGCTGGTTGCCA 548
67 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle..... 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
549 TCCACAGACAGTGGATTCTCTCCATTCGCACTGCTTTGTAGAGAGG 598
81 ..... 81
599 CTGTTCTACTATGAGAGGTAGATCAAGTACAGAGCTTCTGCACAAG 648
82 .....GlyHis..... 83
649 TTAGACTTTATGCTCGCTGCTGCTCATATTTGATGGCTACATCTACAC 698
84 .....AsnArgMetTrpArgIysAsnArgSerPheTyrAlaAsn 97
699 CTGACGACCAAGCCGATTTTGGAGAAACACGCTCCACCCCTACTGGAT 748
97 snHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerIleHisTrp 113
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
749 CTAGCTGCATTGGCCACAGACCCCAACAAATTT...GATCGCTGGTGG 795
114 CysGluGluGlyAlaSerSerSerSerSerGluThrTyrCysGlyLe 130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
796 TGTGAATTGGAGCTCTCGAAACCCCTGTGATGAACTTACTGTGGACC 845
130 uTyrProGluSerGluProGluValIleAlaValAlaSerPheLeuArg 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
846 TGGCGCAGAGCTGTGAAGAAGACCAAGCCCTGGCTGATTTCATCCCA 895
147 rGAsnIleAsnGlnIleIysAlaTyrIleSerMetHisSerTyrSerGln 163
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
896 ACAAACTCTCTCCATCAAGGACATATCTGACAATCCACTCGTACTCCCA 945
164 HisIleValPheProTyrSerTyrThrArgSerIysSerIysAspHisGln 180
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 ATGATGATCTACCTTACTCATATGCTTACAAACTCGGTGAGAACAAATGC 995
180 uGluLeuSerLeuValAlaSerGluAlaValAlaArgAlaIleGluIysTrs 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
996 TGAGTTGAATCCCTGGCTAAAGCTACGTGAAGAACTT...GCCCTAC 1042
197 eRlyAsnThrArgTyrThrHisIleGlyHisGlySerGluThrLeuTyrLeu 213
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1043 TGCACGGACCAAGTACACATATGCGCCGAGAGCTACAAACAATCATATCT 1092
214 AlaProGlyIysGlyIysAspTrpIleTyrAspLeuGlyIleIysTyrSe 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1093 GCTCTGGGGGCTCTGACGACTGGGCTTATGACCAAGAAATCAATATTC 1142
230 rPheThr 232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1143 CTCACCC 1149
seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-860-882A-74
seq_documentation_block:
; Sequence 74, Application US/08860882A
; Patent No. 5985281

```

```

; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGLESTON, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1059 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-860-882A-74

alignment_scores:
Quality: 560.00 Length: 269
Ratio: 3.111 Gaps: 5
Percent Similarity: 66.914 Percent Identity: 40.892

alignment_block:
US-09-980-881-4 x US-08-860-882A-74 ..
Align seg 1/1 to: US-08-860-882A-74 from: 1 to: 1059

1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 GCACTGCTACTCTTACGACAGATACAAAGTGGGAACATAGAGGC 116
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrIleHis 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TTGGACTCAACAAGTCGCCACTGAGAAATCCACCCCTCATCTCCGAGTG 166
34 IsileGlySerSerPheGluTyrTyrProLeuTyrValLeuIysValSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTATCGAACCACATTTGAGGAGCGGCTATTACTCTCGAAGGT... 213
51 GlyIysGluGlnThrAlaIysAsnAlaIleTrpIleAspCysGlyIleH 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GGCAAAGCTGACAAATTAAGCCCTCATTTTCATGAGCTGGTTGCCA 263
67 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle..... 81

```


OM of: US-09-980-881-4 to: EST:* out_format : pfs
Date: Sep 18, 2002 5:41 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+ .p2n.model -DEV-x1h
-O/cg12_1/USPRO.spool/US09980881/runat_16092002_140128_10623/app-query.fasta_1.1126
-DB-EST -QPM-fastap -SUFFIX=first -GAPOP=12.000 -GAPEXT=4.000
-MIMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELDP=6.000
-DELEXT=7.000 -START=1 -MATRIX-biosum62 -TRANS-human40.cdi
-LIST=45 -DOCM=LOCAL -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTEXT=pfs -NORM=ext -HEADSIZE=500
-MILEN=0 -MAXLEN=200000000 -USER=US09980881 @CGNL 1.3434
-NCPU=6 -ICPU=3 -LONGEST -DEV_TIME=OUT=120 -WARN_TIME=OUT=30
-NO_XLXRY -WAIT -THREADS=1

Search information block:

Query: US-09-980-881-4

Query length: 246

Database: EST.*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 3115.820000

score_list:	Strd Orig	ZScore	EScore	len	Documentation
gb_hnc:AK004045	+ 1112.00	2190.10	9.6e-113	1446	I AK004045 Mus musculus 18 days
gb_est2:BG569281	+ 978.00	1928.94	3.4e-98	852	I BG569281 602588573p1 NIH_MGC_76
gb_est1:AV6511709	+ 831.50	1639.14	4.7e-82	688	I AV6511709 AV6511709 GIC Homo sapi
gb_est2:BG567264	+ 828.50	1631.29	1.3e-81	806	I BG567264 602589745p1 NIH_MGC_76
gb_est1:AL536068	+ 812.50	1603.33	4.6e-80	575	I AL536068 AL536068 LTI_FLO13_FBI
gb_est1:AW413090	+ 784.50	1542.33	1.2e-76	889	I AW413090 uq52c03.x1 Sugano mus
gb_est1:AI259599	+ 698.50	1373.17	3.1e-67	723	I AI259599 uq52c03.x1 Sugano mus
gb_est1:BF384332	+ 685.50	1346.46	9.4e-66	772	I BF384332 602584670p1 NCI_CGAP_1
gb_est1:AI253337	+ 659.50	1329.96	8.0e-65	692	I AI253337 u194a04.x1 Sugano mus
gb_est2:BI262107	+ 656.50	1289.33	1.4e-62	713	I BI262107 602958724p1 NCI_CGAP_1
gb_est2:BG642020	+ 649.00	1271.74	8.2e-62	645	I BG642020 p911c.pk002.n16 Chikie
gb_est2:BG61865	+ 631.00	1241.78	4.4e-60	546	I BG61865 yf02b04.r1 Soares fetal
gb_est1:AI595707	+ 624.00	1235.96	1.3e-59	756	I AI595707 u168d08.x1 Sugano mus
gb_est1:BG565100	+ 620.00	1224.13	6.1e-59	747	I BG565100 602583720p1 NIH_MGC_76
gb_est1:AV6511204	+ 607.00	1198.22	1.7e-57	378	I AV6511204 AV6511204 GIC Homo sapi
gb_hnc:AK006678	+ 607.00	1186.30	7.5e-57	1041	I AK006678 Mus musculus adult ma
gb_hnc:AK003061	+ 577.50	1123.72	2.4e-53	1429	I AK003061 Mus musculus adult ma
gb_est2:BG216556	+ 561.50	1098.00	6.5e-52	843	I BG216556 R6736106 Athysys RAGH
gb_est1:AB030473	+ 560.50	1093.08	1.2e-51	1176	I AB030473 AB030423 Mus musculus
gb_est1:AA244760	+ 547.00	1072.97	1.7e-50	616	I AA244760 mx05d08.r1 Soares mus
gb_est2:BI064948	+ 542.00	1056.82	1.3e-49	1022	I BI064948 p911c.pk003.a6 normal1
gb_est2:BF236333	+ 528.00	1034.71	2.2e-48	623	I BF236333 602905932p1 NCI_CGAP_1
gb_est1:AB030424	+ 523.00	1017.78	1.7e-47	1125	I AB030424 AB030424 Mus musculus
gb_est1:AI048167	+ 512.50	1002.01	1.4e-46	724	I AI048167 uq71b12.y1 Sugano mus
gb_est1:AA968389	+ 506.50	991.42	5.6e-46	644	I AA968389 uc71b07.x1 Sugano mus
gb_est1:AV186705	+ 504.50	988.56	8.1e-46	585	I AV186705 u156a06.x1 Sugano mus
gb_est1:AV682330	+ 483.50	941.13	3.6e-43	384	I AV682330 AV682320 GKB Homo sapi
gb_est1:AV652936	+ 481.50	947.17	1.5e-43	379	I AV652936 AL569396 LTI_FLO02_PLI
gb_est1:AV652344	+ 473.50	930.74	1.4e-42	1361	I AV652344 AV652344 GIC Homo sapi
gb_hnc:AK003088	+ 470.50	912.44	1.4e-41	1966	I AK003088 Mus musculus adult ma
gb_est2:BG616436	+ 466.50	910.00	1.9e-41	738	I BG616436 60261618p1 NIH_MGC_76
gb_est1:AI196759	+ 464.50	909.13	1.7e-41	566	I AI196759 u166b06.x1 Sugano mus
gb_est1:AI527277	+ 461.50	910.36	1.9e-41	788	I AI527277 u166b06.x1 Sugano mus
gb_est2:BG618629	+ 456.00	899.26	8.7e-41	725	I BG618629 602645967p1 NIH_MGC_76

gb_est2:BG566051 + 445.00 866.96 4.8e-39 747 I BG566051 602582552p1 NIH_MGC
gb_est2:W89198 - 438.50 856.70 1.8e-38 593 I W89198 z169f104.s1 Soares.fet
gb_est2:BI766188 + 433.00 843.14 1.0e-37 739 I BI766188 603052820p1 NIH_MGC
gb_est2:BI792659 + 431.00 841.64 1.2e-37 598 I BI792659 1c39c12.y1 Melton N
gb_est2:BI712255 + 430.00 840.00 1.5e-37 580 I BI712255 1b77e07.y4 Amplifite

seq_name: gb_hnc:AK004045

seq_documentation_block:

LOCUS AK004045 1446 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:1110032P04:carboxypeptidase B2 (plasma),
full insert sequence.

ACCESSION AK004045 GI:12835067
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
clone:11b:RIKEN full-length enriched mouse cDNA library
clone:1110032P04.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
Carlincl, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

REFERENCE

Carlincl, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carlincl, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashito, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., and Hayashizaki, Y.
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, T., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE

Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

REFERENCE

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arawaka, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
Carlincl, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanaoka, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Hoti, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawal, J., Kojima, Y., Konda, M., Koyu, S.,
Kurahara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numata, K., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schiml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tasaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaoka, T.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

REFERENCE

Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Evolution Research Group, RIKEN Genomic Sciences Center (GSC),

TITLE

Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Evolution Research Group, RIKEN Genomic Sciences Center (GSC),

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 852)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-riemail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: UICM1580 row: b column: 19
High quality sequence stop: 751.
Location/Qualifiers
1..852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4722354"
/clone_lib="NIH-MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Liver; Vector: pDMR-LIB (Clontech); Site_1:
SfiI (ggcgccctcgcc); Site_2: SfiI (ggccatagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-ATCTAGAGCGCGAGCGCCGACATG-dT(30)-BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC library."

BASE COUNT 251 a 177 c 187 g 237 t
ORIGIN

alignment_scores:
Quality: 978.00 Length: 260
Ratio: 4.866 Gaps: 5
Percent Similarity: 77.308 Percent Identity: 75.385

alignment_block:
US-09-980-881-4 x BG569281 ..

Align seg 1/1 to: BG569281 from: 1 to: 852

```

43  ProluTyValleuLysValSerGlyLysGluGlnThraAlaLysAsnAl 59
|||||
2  CCACTCTATGTTTAAAGGTTTCTGGAAAAGAACACACACCAAAATATGC 51
|||||
59  alleTPrleaspcysGlylleHsAlaargLutPrpIleSerProAlar 76
|||||
52  CAATGATGACTGTGGAAATCCATGCCAAGAATGATCTCTCCCTT 101
|||||
76  heCysleuTrpPheIleGlyHis..... 83
|||||
102  TCGGCTTGCGTTCATAGGCAATATACTCAATTCTATGGGATATAGG 151
|||||
83  ..... 83
152  CAATATACCAATCTCGAGGCTTGATTTCTATGCGAGTGT 201
84  .....AsnArgMetTrp Arg 88
202  TAATGTGATGCTTATGACTACTCATGGAAGAAAGATGATGGAGCA 251
89  LysAsnArgSerPheTyAlaAsnAlaCysIleGlyThrAspLeuAs 105
|||||
252  AAGAAGCGTCTTCTATATGGAACAATCATTCATCGGAACAGACCTGA 301

```

```

105  nargAsnPhaIaSerLysHisTrpCysGluGlnGlyAlaSerSerSers 122
|||||
302  TAGGAACCTTCTCTCCAAACACTGGTGTGAGAGAGTGCATCCAGTCTCT 351
|||||
122  erCysSerGluThrTyCysGlyLeuTyrrProGluSerGluProGluVal 138
|||||
352  CATGCTCGGAACCTACTAGTGTACTTTATCCGTGATCGAAGCAAGATGTG 401
|||||
139  LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTy 155
|||||
402  AAGCAGTGCTACTTCTTTCAGAAAGAAATATCAACCGATTAAGCTA 451
|||||
155  rIleSerMetHisSerTyrrSerGlnHisIleValPheProTyrrSerTyrr 172
|||||
452  CATCAGCATGATCATCATCTCCAGCATATAGTGTTCATATTCCTATA 501
|||||
172  hrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
|||||
502  CACGACGTAAGAAAGCAAGACCATGAGGAACGTCTCTAGTACGACGTGA 551
|||||
189  AlAlaValArgAlaIleGlyLysThrSerLysAsnThrArgTyrrThrIsl 205
|||||
552  GCAATTCGTCTATTCAGAAATTAAGTAAATACAGAGTATACATGTC 601
|||||
205  yHisGlySerGluThrLeuTyrrLeuAlaProGlyGlyLysAspTrpI 222
|||||
602  GCATGGCTCAGAAACCTATACCTAGCTCTGAGAGTGGGACGATTTGA 651
|||||
222  leTyrrAspLeuGlylleLysTyrrSerPheThr..... 232
|||||
652  TCTATGATTTGGGCGATCAGATATTCGTTACAAATTCGAACCTTCAGATA 701
|||||
233  .....SerAsn..ProProVa 237
702  CGGCGACATACGATTCCTCTGCCGAGCGTTCATCAACCAACCACTGT 751
|||||
237  LglLysLeuLeuProLeuSerLeu 245
|||||
752  AGAGAGCTTTGGCCGCTCTCTA 776

```

seq_name: gb_est1:AV651709

seq_documentation_block:
LOCUS AV651709 688 bp mRNA linear EST 15-JAN-2002
DEFINITION AV651709 GLC Homo sapiens cDNA clone GLCSEF10 3', mRNA sequence.
ACCESSION AV651709
VERSION AV651709.1 GI:98727223
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 688)
AUTHORS Xu,X., Huang,D., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinoma genesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
JOURNAL Contact: Zeguang Han
MEDLINE Chinese National Human Genome Center at Shanghai
COMMENT 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1..688
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="GLCSP10"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      208 a      145 c      151 g      184 t
ORIGIN

Alignment_scores:
  Quality: 831.50      Length: 181
  Ratio: 5.133      Gaps: 2
  Percent Similarity: 89.503      Percent Identity: 88.950

Alignment_block:
US-09-980-881-4 x AV651709 ..

Align seg 1/1 to: AV651709 from: 1 to: 688

84  AsnATGmetTPArgLysAsnATGSerPheTYrAlaAsnAsnHicys11 100
    |||||||
46  AATCGAATGTGGAGAAAGAACGCTTCTTATGGAGAACATCATTCAT 95
    |||||||
100 eGlyThrAspLeuAsnATGAsnPhaAlaSerLysHis1TrpCysGluGlu 117
    |||||||
96  CGGAACAGACCTGAAATAGCACTTGTCTCCAAACAGCTGCTGAGAAAG 145
    |||||||
117  LysAlaSerSerSerCysSerGluThrTYrCysGlyLeuTYrProGlu 133
    |||||||
146  CTGCATCCAGTTCCTCATGCTCGGAAACCTAGCTGAGACTTATCCTGAG 195
    |||||||
134  SerGluProGluValLysAlaValAlaSerPheLeuATGATGAsn1leas 150
    |||||||
196  TCAGAACAGAGTAGAGAGCAGTGTCTGTTCTTGAGAGAAATATCA 245
    |||||||
150  nGlnIleLysAlaTYrIleSerMetHisSerTYrSerGlnHisIleValP 167
    |||||||
246  CCAGATTAAACATACATACATGATTCATTCATCCAGCATATATAGTGT 295
    |||||||
167  heProTYrSerTYrThrArgSerLysSerLysAspHisGlnGluLeuSer 183
    |||||||
296  TTCCATATTCCTATACAGAAATGAAACCAAGACCATGAGCACTGTCT 345
    |||||||
184  LeuValAlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsnTh 200
    |||||||
346  CTAGTAGCCAGTAGAAGCAGTTCGTCTATGTGAAAACTAGTAAATAATAC 395
    |||||||
200  rArgTYrThrHisGlyHisGlySerGluThrLeuTYrLeuAlaProGlyG 217
    |||||||
396  CAGGTATACATGCGCCATGGCTCAGAAACCTTATACCTAGCTCTGGAG 445
    |||||||
217  LysLysAspAspTrpIleTYrAspLeuGlyIleLysTYrSerPhe..... 231
    |||||||
446  CTGGGAGAGATTGCATATGATTTGGGCATCAAAATATTCGTTTCAATTT 495
    |||||||
232  .....Thrs 233
    |||
496  GAACCTCGAGATACCGGACATACCGATTCTTTGGGCGGAGCGTTACAT 545
    |||
233  eRAsnProProValAGLulLysLeuLeuProLeuSerLeuLys 246
    |||||||
546  CAACCCACACCTGTTGAGAAAGCTTTGGCGCTGTCTTAAA 586

seq_name: gb_est2:BG567264

seq_documentation_block:
LOCUS      BG567264      806 bp      mRNA      linear      EST 10-Apr-2001
DEFINITION  602389745F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723690 5',
            mRNA sequence.
ACCESSION  BG567264

```

```

VERSION      BG567264.1      GI:13574917
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE        1 (bases 1 to 806)
JOURNAL      NIH-MGC http://mgc.ncl.nih.gov/.
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cgabs-remail.nih.gov
              Tissue Procurement: CLONETECH Laboratories, Inc.
              CDNA Library Preparation: CLONETECH Laboratories, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: L1CM1583 row: j column: 11
              High quality sequence stop: 772.
              Location/Qualifiers
                1..806
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4723690"
                /clone_1lb="NIH_MGC_76"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                SfiI (ggccgctggcc); Site_2: SfiI (ggccattatggcc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CAGGCGCATATGACC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGCGCCGAGCGCCGACATG-dT(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size 1.85
                kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
                by PCR. This library was enriched for full-length clones
                and was constructed by Clontech Laboratories (Palo Alto,
                CA). Note: this is a NIH-MGC Library."
BASE COUNT      239 a      158 c      162 g      247 t
ORIGIN

Alignment_scores:
  Quality: 828.50      Length: 179
  Ratio: 5.146      Gaps: 2
  Percent Similarity: 89.944      Percent Identity: 89.385

Alignment_block:
US-09-980-881-4 x BG567264 ..

Align seg 1/1 to: BG567264 from: 1 to: 806

86  MetTPArgLysAsnATGSerPheTYrAlaAsnAsnHicys11eGlyTh 102
    |||||||
2  ATGTGAGAAAGAACCGTCTTCTTATGCGAAACATTCATTCATGCGAAC 51
    |||||||
102 rAspLeuAsnATGAsnPhaAlaSerLysHis1TrpCysGluGluLysAlas 119
    |||||||
52  AGACCTGAATAGCACTTGTCTCCAAACACTGCTGAGAGAGGTGCAT 101
    |||||||
119  eRserSerSerCysSerGluThrTYrCysGlyLeuTYrProGluSerL 135
    |||||||
102  CCAGTTCCTCATGCTCGGAAACCTAGCTGAGCTTATCTGAGCAGAA 151
    |||||||
136  ProGluValLysAlaValAlaSerPheLeuATGATGAsn1leasGln1 152
    |||||||
152  CCAGAGTGAAGCGAGTGTAGTTCTTGTGAGAGAAATATCAACAGAT 201
    |||||||
152  eLysAlaTYrIleSerMetHisSerTYrSerGlnHisIleValPhePro 169
    |||||||
202  TAAAGCATATACATGATCATTCATCTCCAGCATATATAGTTTCAT 251
    |||||||
169  yrSerTYrThrArgSerLysSerLysAspHisGluGluLeuSerLeuVal 185
    |||||||

```

```

|||||
253 ATTCTATACAGAACTAAAGCAAGACCAATGAGAACTGCTCTAGTA 301
186 AAlaSerGluAlaValAlaAlaIleGluTyrSerLysAsnThrArgTy 202
302 GCCAGTGAAGCAGTTGCTGCTATTGAGAAACAGTAAATAATCCAGTA 351
202 rThrsHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAla 219
352 TACACATGGCCATGGCTCAGAAACCTATACCTAGCTCTGGAGTGGGG 401
219 sPASPTrIleTyrAspLeuGlyIleLysTyrSer.Phe..... 231
402 ACGATTGGATCTATGATTGGCATCAAAATATCGTTACAAATTGACTT 451
232 .....ThrsAsp 235
452 CGAGATACGGGCAATACGATTCTTGCCTGGAGCGCTTACTCAAC 501
235 rProValGluLysLeuLeuProLeuSerLeuLys 246
502 CACCTGTAGAGAGCTTTGCCGCTGCTCTATAA 536
seq_name: gb_est1:AL536068
seq_documentation_block:
LOCUS AL536068 575 bp mRNA linear EST 13-FEB-2001
DEFINITION AL536068 LTI_FL013_FBRn1 Homo sapiens cDNA clone CS0DF022YE21 3
prime, mRNA sequence.
ACCESSION AL536068
VERSION AL536068.1 GI:12799561
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 575)
L1.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
Location/Qualifiers
1..575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
Maryland 20850, USA Fax : (1) 301 610 8371 Email :
filiang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 151 a 129 c 125 g 168 t 2 others
ORIGIN
alignment_scores:
Quality: 812.50 Length: 182
Ratio: 5.015 Gaps: 3
Percent Similarity: 89.011 Percent Identity: 87.363
alignment_block:

```

```

US-09-980-881-4 x AL536068/rev ..
Align seq 1/1 to reverse of: AL536068 from: 1 to: 575
84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysI 100
556 AATCGAATGTGGAGAAAGAACCGTTCTTCTATGCCAACAATCTGCAT 507
100 eGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGlu 117
506 TGGACACAGACCTGATATGAACTTTGCTTCCAAACACTGGTGTAGAGAG 457
117 LysAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 133
456 GTGCATCCAGTCTCTCATGCTCGAAACCTACTGTCGACTTATCCCTGAG 407
134 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnLeuA 150
406 TCAGAACCAAGAGTGAAGGCGAGTGGCTAGTTTCTTGAGAAAGAAATCA 357
150 sngInIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 166
356 GCCAGATTAAAGCATATACATCAGCATGATTCATATACCCAGCATATAGTG 307
167 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeu 183
306 TTTCATATTCTCTATACACGAACTAAAGCAAGACCATGAGAACTGTC 257
183 rLeuValAlaSerGluValAlaValAlaIleGluLysThrSerLysAsnT 200
256 TCTAGTAGCCAGTGAAGACAGTCTGCTATTGAGAAACAGTAAATA 207
200 hrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGly 216
206 CCAGATATACATAGCGCATGGCTCAGAAACCTATACCTAGCTCCGGA 157
217 GlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr.. 232
156 GGTGGGACCATTTGATCTGATTTGGCATCAATTTTCGTTTCAAT 107
233 .....S 233
106 TGAACTTCGAGTTACGGGCACATACGATTCTTGCTGCCGGGCGTTTCC 57
233 eRAsn.ProProValGluLysLeuProLeuSerLeuLys 246
56 TCAACCCCCCTGTAGAGAGCTTTGCCGCTGCTCTATAA 15
seq_name: gb_est1:AL536069
seq_documentation_block:
LOCUS AL536069 889 bp mRNA linear EST 13-FEB-2001
DEFINITION AL536069 LTI_FL013_FBRn1 Homo sapiens cDNA clone CS0DF022YE21 5
prime, mRNA sequence.
ACCESSION AL536069
VERSION AL536069.1 GI:12799562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 889)
L1.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
Location/Qualifiers
1..889
/organism="Homo sapiens"
/db_xref="taxon:9606"

```


84 AsnaagmetrPrarglsasnarSerPheTyralaasnaHiscys11 100
||||| ||||||| ||||| :||| :||||| :
665 AATCGAAAGTGAAGCAGACAGCCCTCTGCTCACAAGACACACCCCTGCT 636
100 eglYthrAspleuAanArgAsnPheAlaSerDyShs1rPcysgluglug 117
:||||| :||||| :||||| :||||| :||||| :
635 GGGCAGACAGCTTGAACAGSAACTTGCTCCCAACACCTGCTGAGAAAG 586
117 lylAlaSerSerSerSerCysSergluThrTyrCysGlyLeuTyrProglu 133
||||| ||||||| ||||||| ||||||| ||||||| ||
585 GGGCGTCAAGTTCCTGCTGCTCTGAAACCTACTGTGGACTTATCTGTGAG 536
134 SergluProgluVallylsalAlaValAlaSerPheLeuArgArgAsn1leas 150
||||| ||||||| ||||||| ||||||| ||||||| :
535 TCTGAGCCAGAGCTGAGGCAGCTGCTGCTCTTGAGAGAAATATTCGA 486
150 nglnllyAlaTyrIleSerMetHisSerTyrSerGlnHisIleValp 167
:||||| :||||| :||||| :||||| :||||| :
485 CCACATTAAAGCTTACATCAGTATGCATCATCACTCCCAACAAATACCTGT 436
167 heProTyrSerTyrThrArgSerlySerlyAspHisIleuGluleuSer 183
||||| ||||||| ||||||| ||||||| ||||||| ||
435 TTCCCTATTCCTATACAGACAGCAAAAGACAGACACAGCAAGAACTGT 386
184 lcuValAlaSerGluAlaValAlaArgAlaIleGluTyrShSerlysanTh 200
||||| ||||||| ||||||| ||||||| ||||||| ||
385 CTAGTGGCCAGCGAGCAAGCACTTGCTGCAATTCGAAGATTAATTAACAC 336
200 rArgTyrThrHisGlyHisGlySerGluThrLeuTyrIleuAlaProgluY 217
||||| ||||||| ||||||| ||||||| ||||||| ||
335 CAGGTACACACACAGCGAGTGGCTCAGAAAGTTATATCTAGCTCTGGAG 286
217 lylGlyAspAspThrPrlIeTyrAspLeuGlyIleLeysTyrSer.Phe.... 231
||:::||||| ||||||| ||||||| ||||||| :||
285 GTTCTGACGATTGATGATCATGATTGGGATCATCAAAATATTCGTTACAATT 236
232Thrs 233
235 GAGCTCCGAGATACAGGACAGATACGATTCTTGCTCCCTGAGAGATACATT 186
233 eTAsnProProValGlnTylsLeuLeuProLeuSerLeuLys 246
||||| ||:::||||| || |||||||
185 CAACCCACTTGTCGAGAAGCTTGCGCCGACCATCTCTAATAA 145
seq_name: gb_est1.A1255929
seq_documentation_block:
LOCUS A1255929 772 bp mRNA linear EST 12-NOV-1998
DEFINITION u183d11.x1 Sugano mouse liver mla Mus musculus cDNA clone
IMAGE1889013 3' similar to TR:015114 Q15114 PREPRO-PLASMA
CARBOXYPEPTIDASE B. ;, mRNA sequence.
ACCESSION A1255929
VERSION A1255929.1 GI:3863454
SOURCE EST.
ORGANISM Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 772)
AUTHORS Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,D., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Thelning,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Matra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LINL; contact the
IMAG Consortium (info@image.llnl.gov) for further information.
MGI:973337
Possible reversed clone; similarity on wrong strand
Seq. primer: custom primer used
High quality sequence stop: 477.

FEATURES

source

```

/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone_image="IMAGE:1889013"
/clone_lib="Sugano mouse liver mla"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"

```

BASE COUNT
ORIGIN

204 a 158 c 167 g 238 t 5 others

(note: *Organ: liver; Vector: pME18S-FL3; Site.1: DraIII /CACCCTGTGG); Site.2: DraIII (CACCAATGTG); 1st strand cDNA was primed with an oligo(dT) primer [AATGGCCCTTTTATTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor (TGTGTGGCCCTACTGTG), digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACCTGTGG, 3' site CACCAATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGTCTTAATAAGCTCGC and 3' end primer CGAACCTGCAGCTCAGACA."

alignment_scores:

Quality:	685.50	Length:	201
Ratio:	4.339	Gaps:	4
Percent Similarity:	78.607	Percent Identity:	69.6522

alignment_block:

US-09-980-881-4 x AI255929/rev

Align seg 1/1 to reverse of: AI255929 from: 1 to: 772

67 HisAla.....ArgGluTrpIleSerProalaphcysLeuTrpPh 80
|||||
|||||:::

761 CATGCCCGTGATGACGTGATGGCTA.....TGACTACA 727

```

80 eileglyhisasnargmettrparglysasnargserphetyralaasna 97
   :: ||||| ||||| ||||| :: ||||

```

726 CGTGGAAAGATCGAATGTGAGAAAGANCCGCTCTGCTCACAGAACA 677

97 sNtHScysIleGlyThrAspLeuAsnArgAsnPhenAlaSerLysHisTrp 113

676 ACCGCTGCGTGGGCACAGACTTGAACAGGACTTCGCTTCCAAACACTGG 627

114 CysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLe 130
|||||:::|||||

626 TGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAACCTACTGTGACT 577

130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 147
|||||

576 TTATCNTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAA 527

```

147 rgasnlleasnglnlelysalatyrleesermethisserTyrSergln 163
|||||:::|||||

```

526 GAAATATCGACCACATTANAGCTTACATCAGTATGCACCTCATCTCCCA 477

```

164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisG1 180
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

476 CAAATACGTTCCTATTCTATAACAGAGCAAAGCAAGGCCGCA 427

180 ugiLueSerLeuValAlaSerGluAlaValArgAlaIleGluLysThr 197

```

|||||
426 AAGACTGCTCTAGTGGCCAGGAGACAGTCTGCAATTGAAAGTATTA 377
197 eRLysAsnThrArqTYrThrHISGLYHISGLYSerGIuThrLeuTYrLeu 213
376 ATMAAAACGACGATACACACAGGAGTGGCTGCAAAAGTTATATCTTA 327
214 AAlaProGLYGLYAspAspTrpIleTYrAspLeuGLYIleLysTYrSe 230
326 GCTCCGTGAGTTCGACGATGATCTATGATTGGCATCAATATATTC 277
230 r.Phe..... 231
276 GTTTCACATTGAGCTCCGAGATACAGGAGATACGAGATTCTGCGCTG 227
232 .....ThSerAsnProProValGIuLysLeuLeuProLeuSerLeu 246
226 AGAGTACATCAAAACCACTTGTGCAAGAGCTTGGCCCATCTCTAA 177
246 s 246
176 A 176
seq_name: gb_est2:BF384322
seq_documentation_block:
LOCUS BF384322 692 bp mRNA linear EST 27-NOV-2000
DEFINITION 602046670F1 NCI_CGAP_119 Mus musculus cDNA clone IMAGE:4196129 5',
mRNA sequence.
ACCESSION BF384322
VERSION BF384322.1 GI:11365627
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.lnl.gov
Plate: L14M9532 row: 1 column: 18
High quality sequence stop: 689.
FEATURES
source
location/Qualifiers
1..692
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4196129"
/lab_host="NCI_CGAP_119"
/note="Organ: liver; Vector: pCMV-SPORT6; Site: 1; Note:
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 211 a 156 c 139 g 186 t
ORIGIN
alignment_scores:
Quality: 676.50 Length: 174
Ratio: 4.340 Gaps: 2
Percent Similarity: 85.632 Percent Identity: 76.437
alignment_block:

```

```

US-09-980-881-4 x BF384322 ..
Align seg 1/1 to: BF384322 from: 1 to: 692
91 ArgSerPheTYrAlaAsnAsnHIScysIleGLYThrAspLeuAsnArgAs 107
|||||
1 CGCTGTGCTCACAGAACGACCGCTGCGGGCAGACAGCTTAACAGGAA 50
107 nPheAlaSerLysSHISrPcysGLuGLuGLYAlaSerSerSerCys 124
51 CTTCCGCTTCCAAACACTGCTGTGAGAAAGCTGCTCAAGTCTCTCTGCT 100
124 eRLuThrTYrCysGLYLeuTYrProGLuSerGLuProGLuValLysAla 140
|||||
101 CTGAACCTACTGTGACTTTATCCGACTGTGAGCCAGCGAAGCGCA 150
141 ValAlaSerPheLeuArgArgAsnIleAsnGLIleLysAlaTYrLese 157
|||||
151 GTGGCTGACTTCTTGAGAAATATCGACACATTAAAGCTTACATCAG 200
157 rMetHISerTYrSerGLuHISIleValPheProTYrSerTYrTrpArg 174
201 TATGCATCATCTATCCCAACAAATACTGTTCTTATCTTCTATTAACAGA 250
174 eRLysSerLysAspHISGLuGLuLeuSerLeuValAlaSerGLuAlaVal 190
|||||
251 GCAAAAGCAAGGACCAAGCAAGTGTCTGTAGTGGCCAGCAAGCAGTT 300
191 ArgAlaIleGLuLysThrSerLysAsnThrArqTYrThrHISGLYHISGL 207
|||||
301 CGTGCAATTGAAAGTATTATTAATAAACACACAGTACACACAGCGAGTGG 350
207 ySerGIuThrLeuTYrLeuAlaProGLYGLYAspAspTrpIleTYr 224
|||||
351 CTCGAAAGTTATATATCTAGCTCCGAGAGTCTCTACGATTGATCTATG 400
224 sPLeuGLYIleLysTYrSer.Phe..... 231
|||||
401 ATTTGGCATCAAAATATTCGTTTCAATGAGCTCCGAGATACAGCAGA 450
232 .....ThSerAsnProProValGIuLysL 240
451 TACGATTTCTTCTGCTCCGAGAGATACATCAAAACCACTGTGCGAAGC 500
240 eULeuProLeuSerLeuLys 246
|||||
501 TTTCGCCCGCATCTCTATAAA 520
seq_name: gb_est1:A1255397
seq_documentation_block:
LOCUS A1255397 747 bp mRNA linear EST 12-NOV-1998
DEFINITION u194a04.x1 Sugano mouse liver mlia Mus musculus cDNA clone
IMAGE:1890030.3', similar to TR:015114 Q15114 PREPRO-PLASMA
CARBOXYPEPTIDASE B.; mRNA sequence.
ACCESSION A1255397
VERSION A1255397.1 GI:3862922
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 747)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicinep

```

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:974354

Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 471.
Location/Qualifiers

FEATURES

source

1..747
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone_image="1890030"
/clone_1lb="Sugano mouse liver ml1a"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-F13; Site_1: DraIII
(CACTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor (TGTGGCCCTACTG), digested
and cloned into distinct DraIII sites of the pME18S-F13
vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTTAACACTGCG and 3' end
primer CGACCTGCACCTCGACACA."

BASE COUNT 199 a 151 c 162 g 233 t 2 others
ORIGIN

alignment_scores:
Quality: 659.50 Length: 181
Ratio: 4.368 Gaps: 2
Percent Similarity: 83.425 Percent Identity: 72.376

alignment_block:
US-09-980-881-4 x AI255397/rev ..

Align seg 1/1 to reverse of: AI255397 from: 1 to: 747

```

84 AsnArgMetIrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
|||||
711 AATCGAATGTGGAGAGACCGCTTGTCTCAACAGACCAACCGTGGNGT 662
|||||
100 eGIYThrAspLeuAsnArgAsnPheAlaSerLysHisIstPCysGluGluG 117
|||||
661 GGGCAGACAGACTGACAGGAACTTCTTCAAACTGGTGTGAGAAAG 612
|||||
117 LysAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 133
|||||
611 GTGCGCAAGTTCCTCTCTCTTAAACCTACTGTGACTTATCTCTGAG 562
|||||
134 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAs 150
|||||
561 TCTGAGCCGAGGTGAGGACAGTGGCTGACTTCTTGAAGAAATATCGA 512
|||||
150 nGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValP 167
|||||
511 CCACATTAAGCTTACATCAGTATGCACTATCTCCCAACAAATCTGTG 462
|||||
167 hepProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
|||||
461 TTCCCATTTCTTATACAGAACAAAGCAGACCCGAAAGAACTGTCT 412
|||||
184 LeuValAlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsn 200
|||||

```

```

411 CTAGTGGCCAGCAGACGAGTTCGTGCAATTGAAGTATTAATAAACAC 362
200 TargTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyG 217
|||||
361 CAGGTACACACAGCGAGCGTGGCTCAGAAAGTTATATCTAGCTCTGGAG 312
|||||
217 LysGlyAspAspTIPILeTyrAspLeuGlyIleLysTyrSer.Phe..... 231
|||||
311 GTTCTGACAGATTGGATCTATGATTGGCAGATCAAAATATGCTTACAAAT 262
|||||
232 .....Thrs 233
261 GAGTCCGAGATACAGCAGATACGAGATTCTTGTCTGCTGAGATACAT 212
|||||
233 eRAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
|||||
211 CAACCCACATTTGTGACAGAGCTTTGGCCGCGATCTCTAAAAA 171
|||||
seq_name: gb.est2:BI246210

```

seq_documentation_block:
LOCUS BI246210 713 bp mRNA linear EST 17-JUL-2001
DEFINITION 602958724F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5124700 5',
mRNA sequence.
ACCESSION BI246210
KEYWORDS BI246210.1 GI:14789700
SOURCE EST.
ORGANISM house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 713)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM11304 row: 0 column: 05
High quality sequence stop: 686.

FEATURES
source
Location/Qualifiers

1..713
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="5124700"
/clone_1lb="NCI_CGAP_L19"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 212 a 168 c 145 g 188 t
ORIGIN

alignment_scores:
Quality: 656.50 Length: 175
Ratio: 4.406 Gaps: 3
Percent Similarity: 85.143 Percent Identity: 74.857

alignment_block:
US-09-980-881-4 x BI246210 ..

Align seg 1/1 to: BI246210 from: 1 to: 713

91 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgA 107


```
seq_name: gb_est2:H61865

Seq_documentation_block:      546 bp      mRNA      linear      EST 06-Oct-1995
LOCUS      H61865
DEFINITION      yf02b04.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:204079 5' similar to SP:A41204 A41204 CARBOXYPEPTIDASE B ; ,
RNA sequence.
ACCESSION      H61865
VERSION      H61865.1 GI:1014697
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 546)
AUTHORS      Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Ricklin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston
            R., Williamson,A., Wohlmann,P. and Wilson,R.
            The WashU-Merck EST Project
TITLE      Unpublished (1995)
JOURNAL      Contact: Wilson RK
COMMENT      Washington University School of Medicine
            444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 973
            High quality sequence stops: 377
            Source: IMAGE Consortium, LNLN.
            This clone is available royalty-free through LNLN ; contact the
            IMAGE Consortium (info@image.lnl.gov) for further information.
            Insert Length: 973 Std Error: 0.00
            Seq primer: M13RPI
            High quality sequence stop: 377.
FEATURES
    source
        location/qualifiers
            1..546
                /organism="Homo sapiens"
                /db_xref="GDB:3773210"
                /db_xref="taxon:9606"
                /clone="IMAGE:204079"
                /clone_lib="Soares fetal liver spleen INFLS"
                /sex="male"
                /dev_stage="20 week-post conception fetus"
                /lab_host="DH10B (ampicillin resistant)"
                /note="Organ: Liver and Spleen; Vector: pTZ19D (Pharmacia)
                    with a modified polylinker; Site.1: Pac I; Site.2: Eco RI adaptors
                    double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Pac I and cloned into the Pac I
                    and Eco RI sites of the modified pTZ19 vector. Library
                    went through one round of normalization. Library
                    constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      148 a      112 c      121 g      159 t      6 others
ORIGIN
alignment_scores:
    Quality:      631.00      Length:      154
    Ratio:         4.780      Gaps:          4
Percent Similarity: 85.714      Percent Identity: 84.416
alignment_block:
US-09-980-881-4 x H61865      ..
Align seg 1/1 to: H61865 from: 1 to: 546
113 TTPCGAGUGLUGLUALAASerSerSerCysSerGIunThrTCySGI 129
|||||
13 TGCTGTGCAGGAAGCTGCATCCACTTCTCATGCTCGGAAMCACTACGTGG 62
129 yfeutyrProGLuSerGlunProGiUnVallysaVaAlaAserPheLeua 146
```

```

seq_name: gb_est1:A1595707
seq_documentation_block:
LOCUS
DEFINITION
IMAGE:12925103 3' similar to TR:Q15114 Q15114 PREPRO-PLASMA
CARBOXYPEPTIDASE B. ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .756

```



```
17 rTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleH 34
|||||
239 TTGGATAGAAATTATTAAGTAGAGCATCTGATAGCTTACAAAAATCC 288
34 IsIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 50
|||||
289 ACATTGGATCCCATTTGAGAACTACCCACTCATGTGTTAAAGGTTTCT 338
51 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleH 67
|||||
339 GGAAAAAGAACAGACAGCCAAAAATGCCATATGATTGACTGTGGAATCCA 388
67 sAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 83
|||||
389 TGCCAGAGAAATGAGCCTCTCTGCTTCTGCTTGCTGATAGGCCATA 438
83 ..... 83
439 TAACTCAAATTCATGGGATATAGGGCAATATACCAATCTCTGAGGCTT 488
84 .....AsnArgMe 86
489 GTGATTTCTATGTATGCCAGTGGTTAATGTGATGTTATGACTACTC 538
|||||
86 tTrpArg.....LysAsnArgSerPheTyrAla 96
|||||
539 ATGGAACACAGAAATCGAATGTGAGAACAGAACCGTTCTTCTATGCGA 588
96 sAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHis 112
|||||
589 ACAATCATTCATCGGACAGACACTGATATAGGAATT.GCTTCCAAACAC 637
113 TrpCysGluGln.GlyAlaSerSerSerSer.CysSerGlu.ThrTyr.C 128
|||||
638 TGGGTGTGAGGAAAGGTGATCCAGTTCCTCATGTCTCGAAAAACCTACTT 687
128 ysgLysLeuTyrProGluSerGluProGluValLysAlaValAlaSerPhe 144
|||||
688 GTGGACTTTATCTCTGAGTCAGAACAGAGTG.GAAGCAGTGGCTAGTTTC 736
145 LeuArg 146
|||||
737 TTGAG 742
```

)

This Page Blank (uspto)